

What efficacy and safety profiles can we expect

*Mathematical and Computational Biology in Drug Discovery
(MCBDD) Module IV*

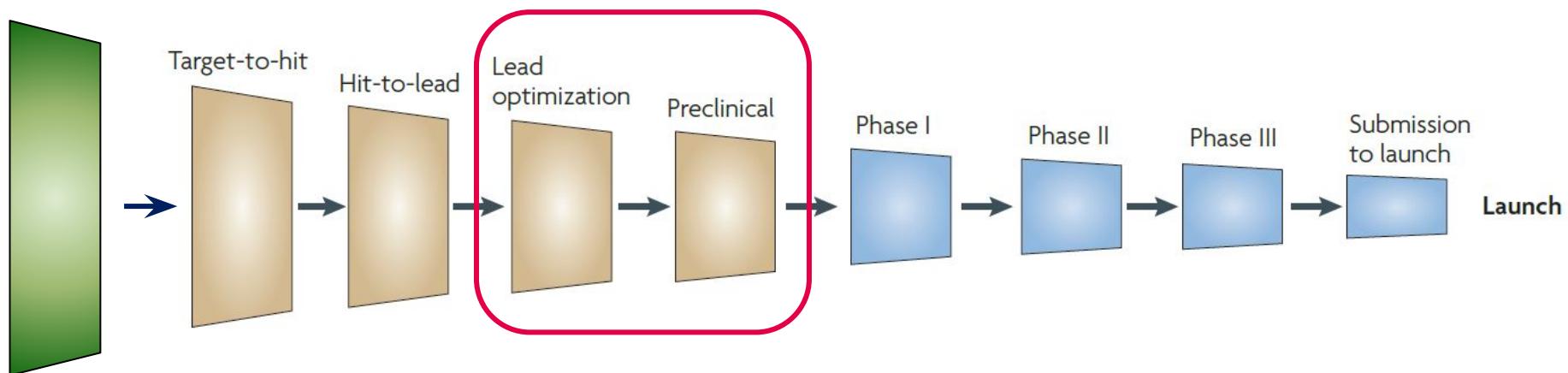
*Dr. Jitao David Zhang
April-May 2021*

The outline of the module

- MoA understanding of the compound with gene expression
 - Single-cell gene expression
 - RNA velocity
 - Spatial transcriptomics
- Target identification
 - CESTA
 - Degradation
 - Binding mode

Where are we now

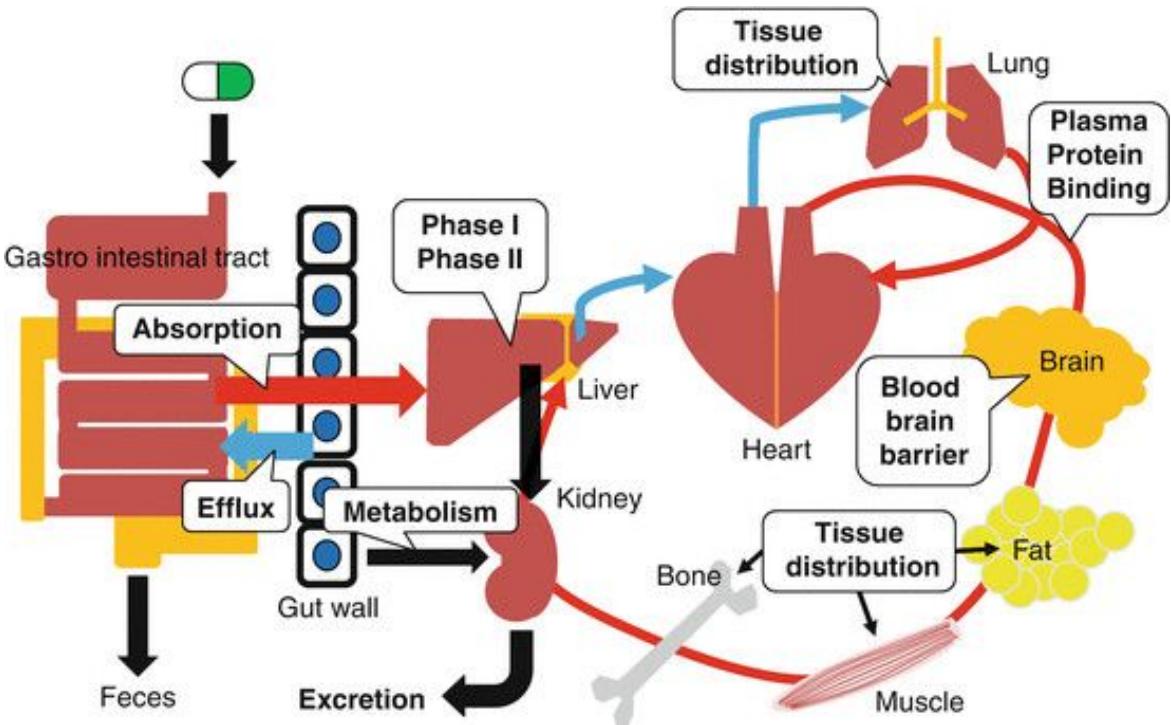
Target identification & assessment



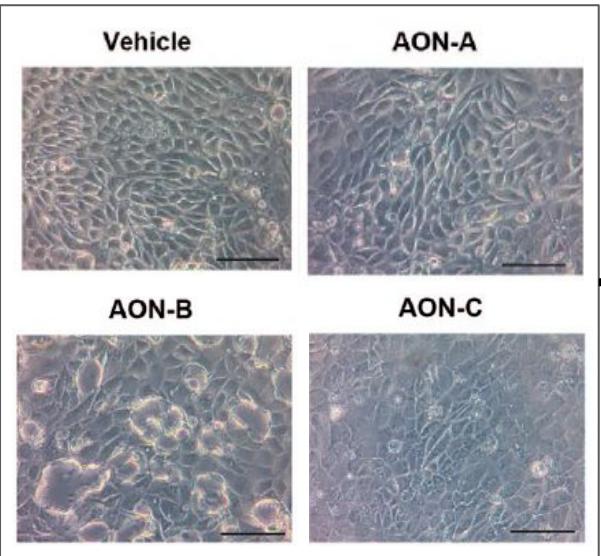
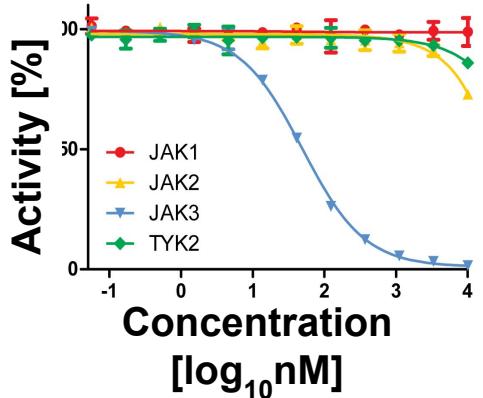
Goal: we want to select **one compound** from a few (~ 10^2 - 10^0) for entry in human.

Factors that affect efficacy and safety profiles

- Absorption
- Distribution
- **Pharmacology**
- **Toxicology**
- Metabolism
- Excretion

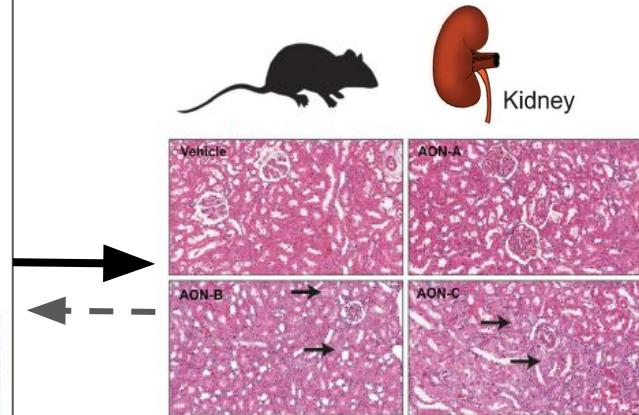


Classical workflow of efficacy and toxicity assessment



Cellular assays
(in vitro)

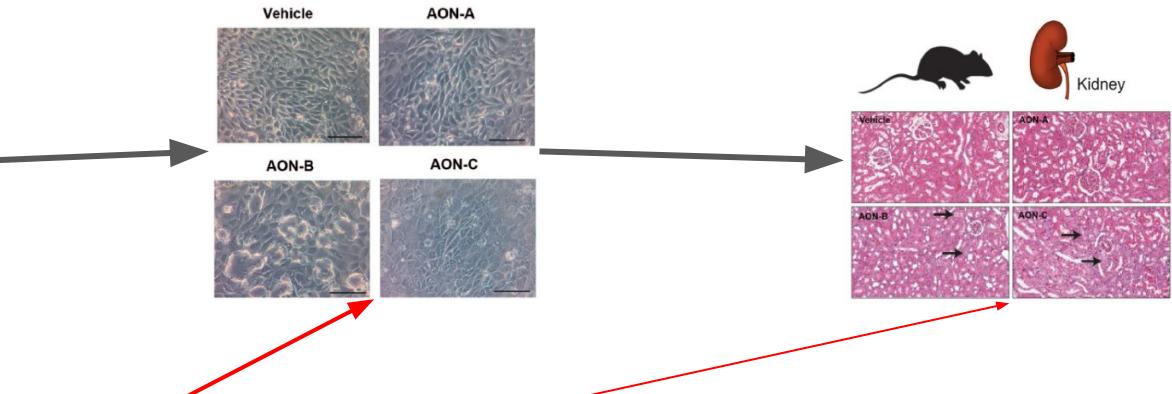
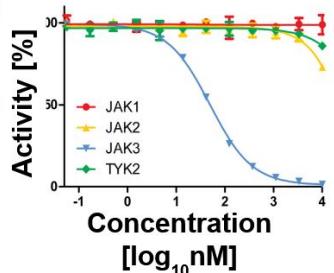
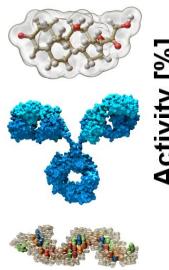
Biochemical & biophysical assays



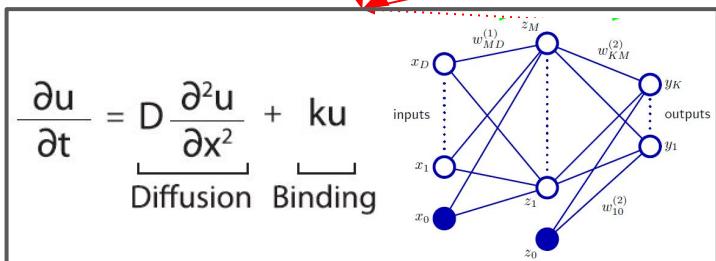
Animal experiments
*(*in vivo*)*

→ Usual workflow
 ← - - Assay development

Computational methods empower efficacy and toxicity assessment



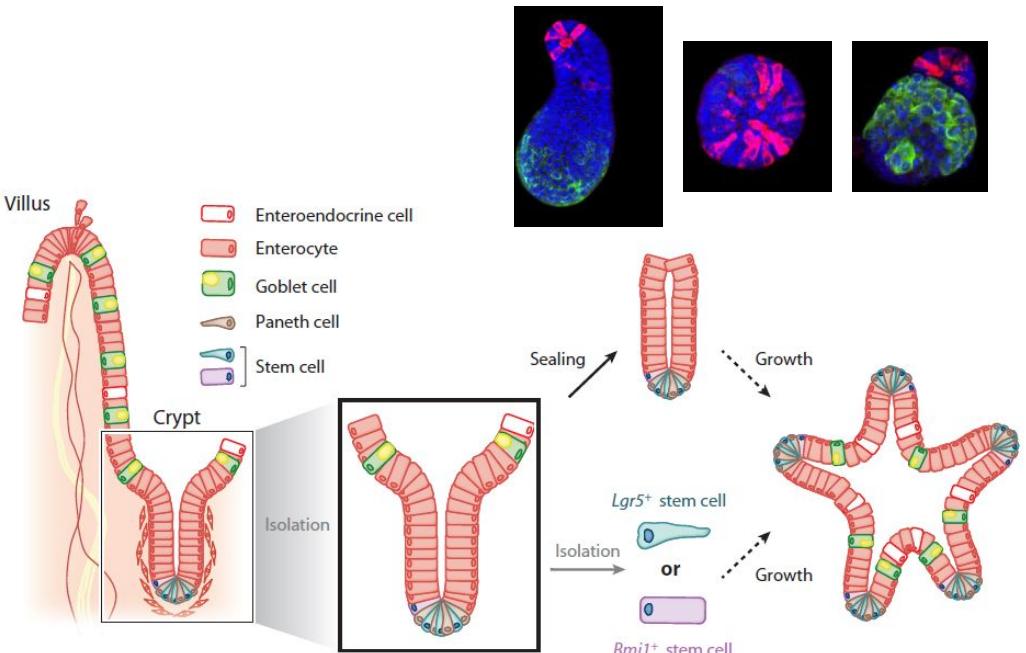
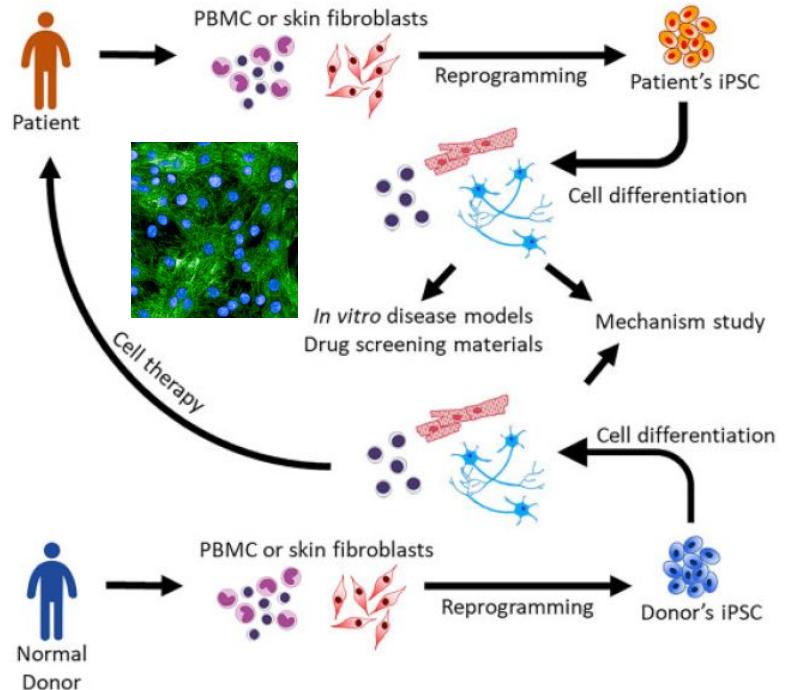
High-throughput technologies (omics, microscopy, etc.)



Mechanistic and statistical models



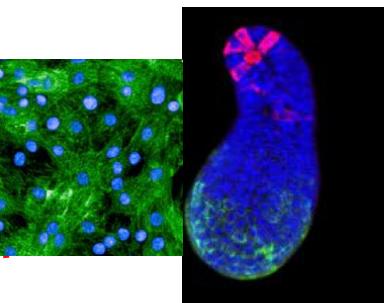
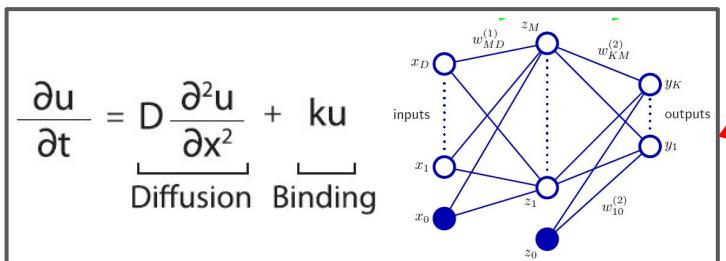
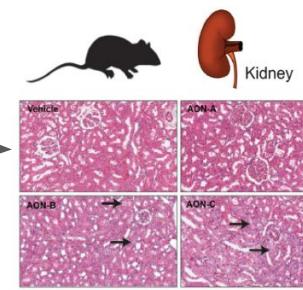
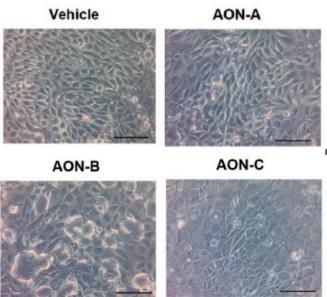
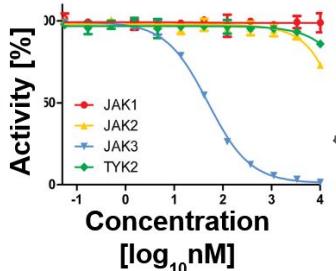
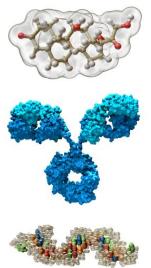
Stem cells and organoids empower efficacy and toxicity assessment



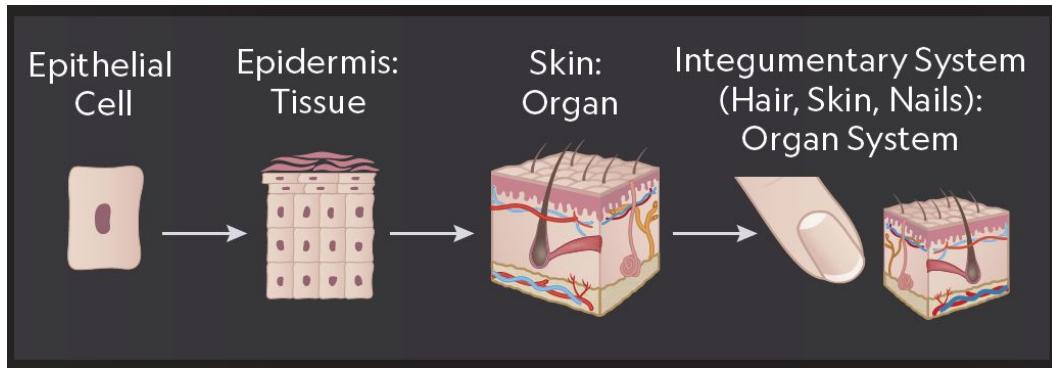
Small-intestinal organoids

Induced pluripotent stem-cells

Computational methods and novel biological models empower efficacy and toxicity assessment



Complexity Increases Through a System

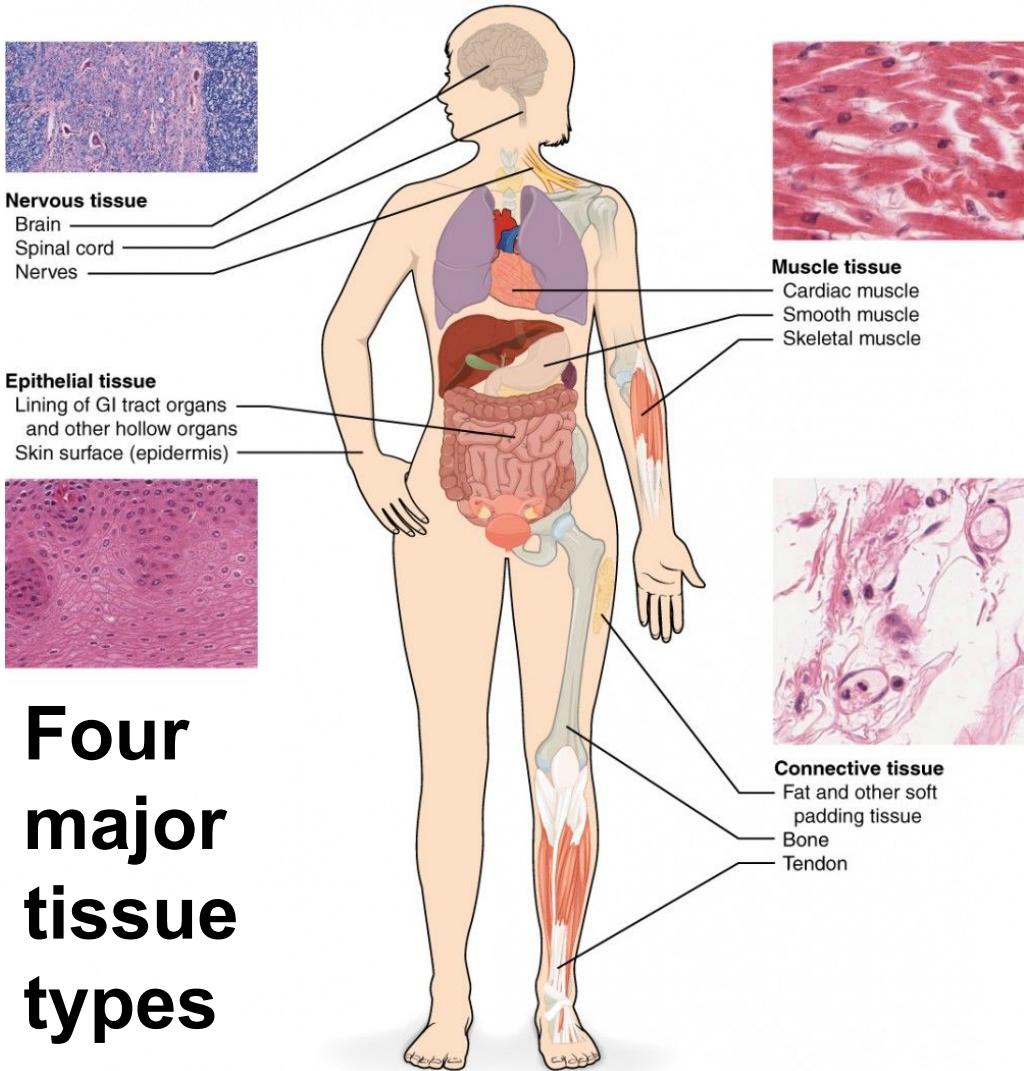
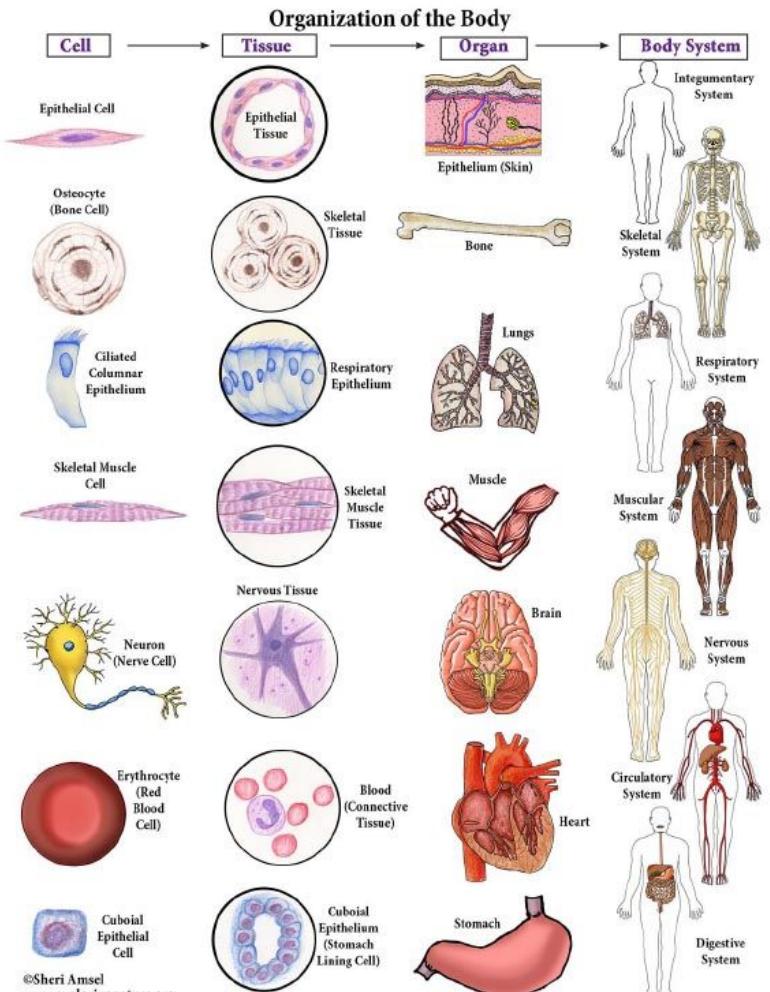


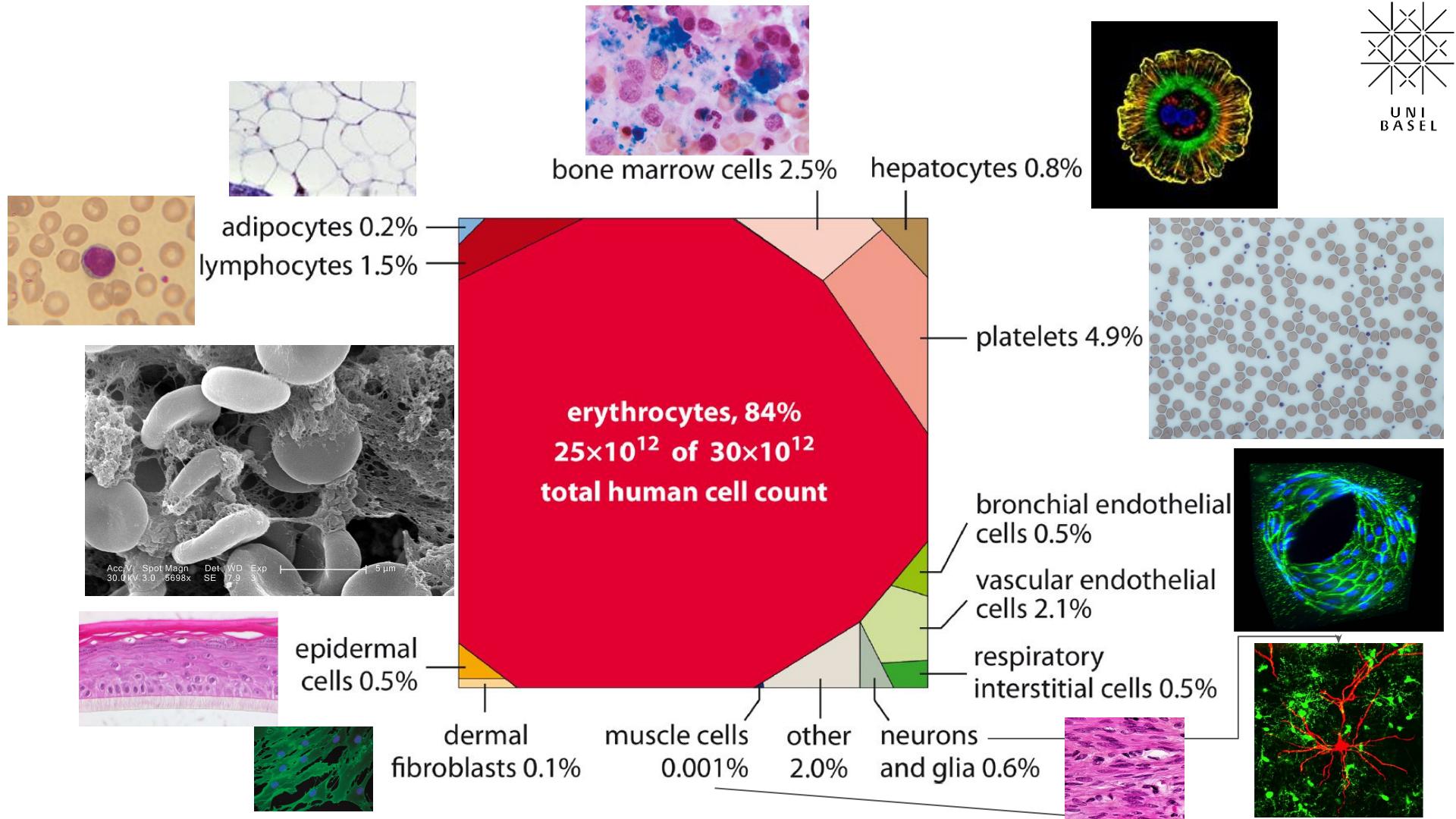
Cells: basic building blocks, of variable morphologies and functions

Tissues: groups of specialized cells that communicate and collaborate

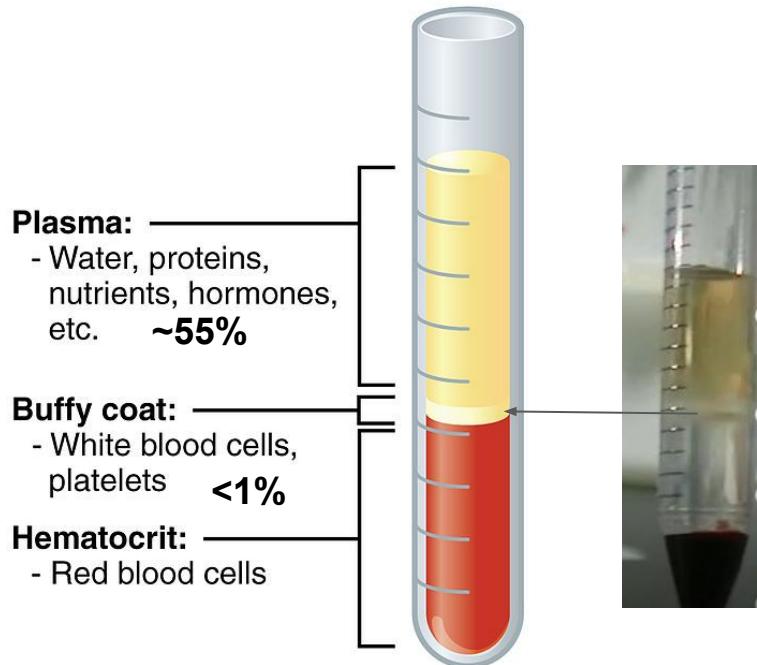
Organ: group of tissues to perform specific functions

Organ systems: group of organs and tissues



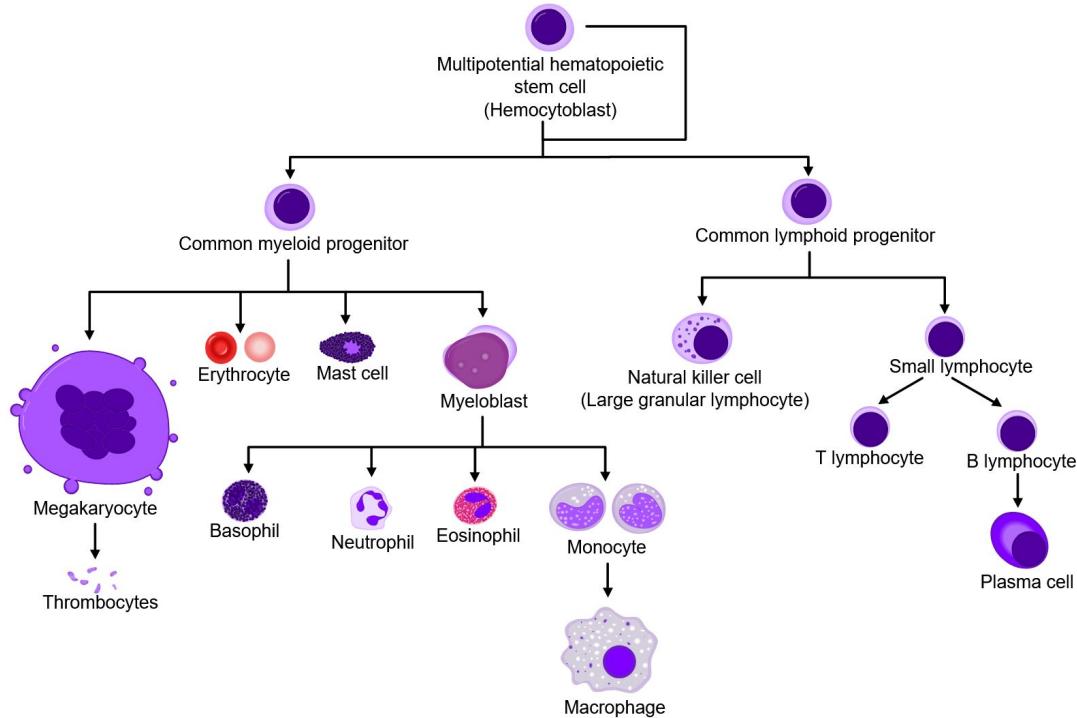


What's in a drop of blood? Ask a doctor!



Normal Blood:

♀ 37%–47% hematocrit
 ♂ 42%–52% hematocrit

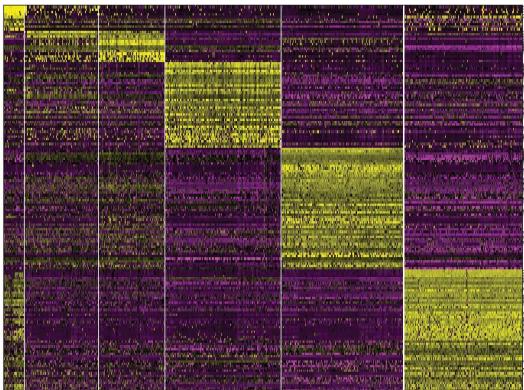


What's in a drop of blood? Count the genes!



Sequencing

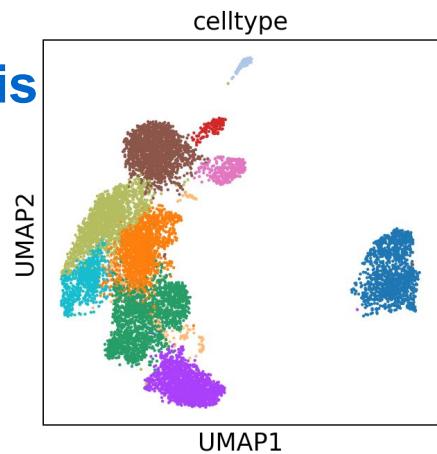
Genes



Cells

Low Expression  High Expression

Data analysis

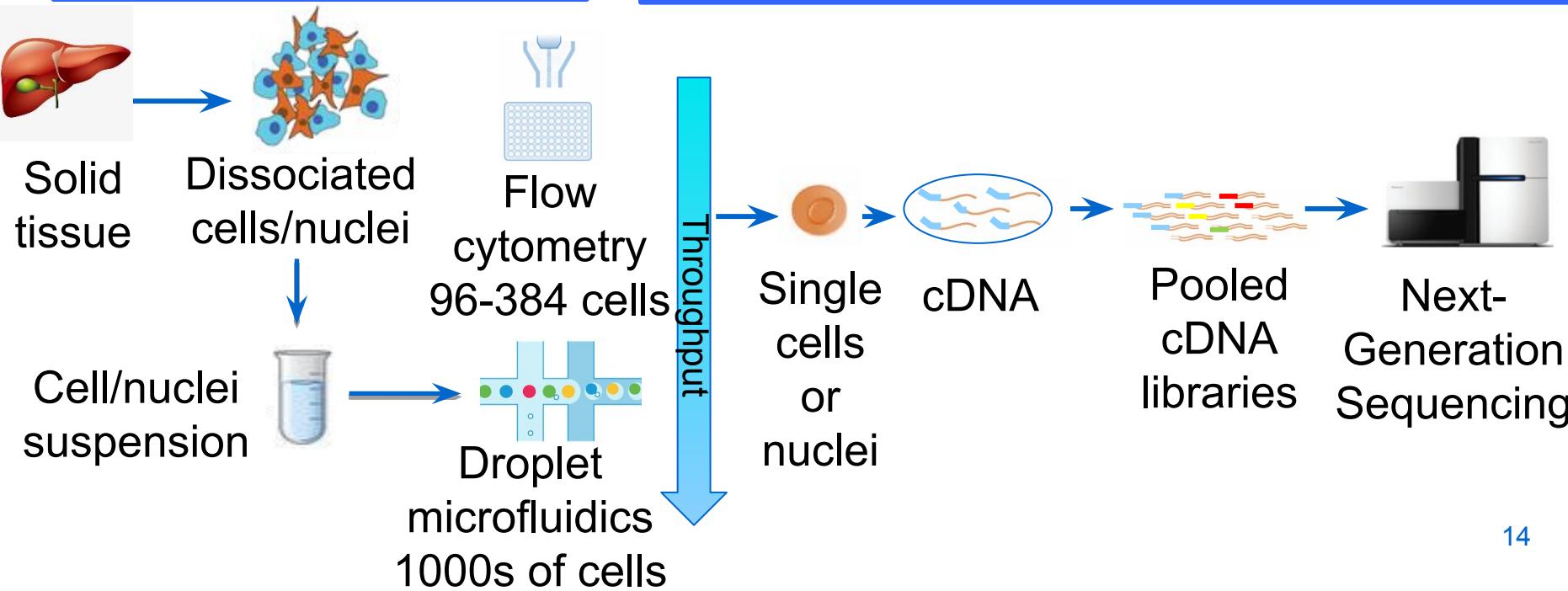


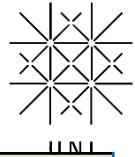
- B-cell
- CD4 T-cell
- CD8 T-cell
- DC
- NK cell
- monocyte CD14+
- monocyte CD16+
- naive CD4 T-cell
- naive CD8 T-cell
- pDC
- unknown

Single-cell sequencing (scSeq) workflow

Tissue dissociation

Single cell capture and transcriptome sequencing





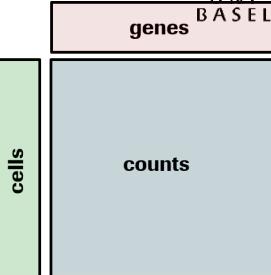
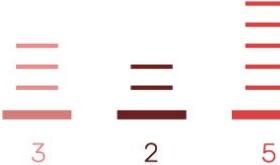
A linearized workflow of scSeq data analysis

From short reads to gene-cell matrix

Alignment

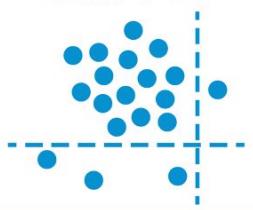


Quantification

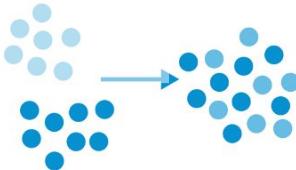


QC, filtering & normalization,
dimensionality reduction, and
clustering

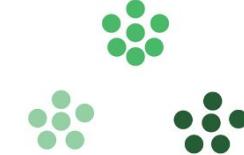
Quality control



Normalisation

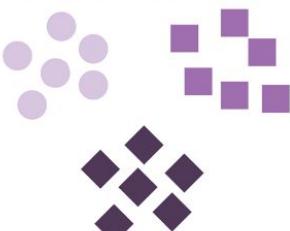


Clustering



Downstream analysis

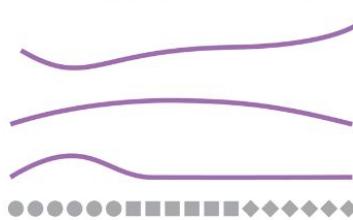
Differential expression



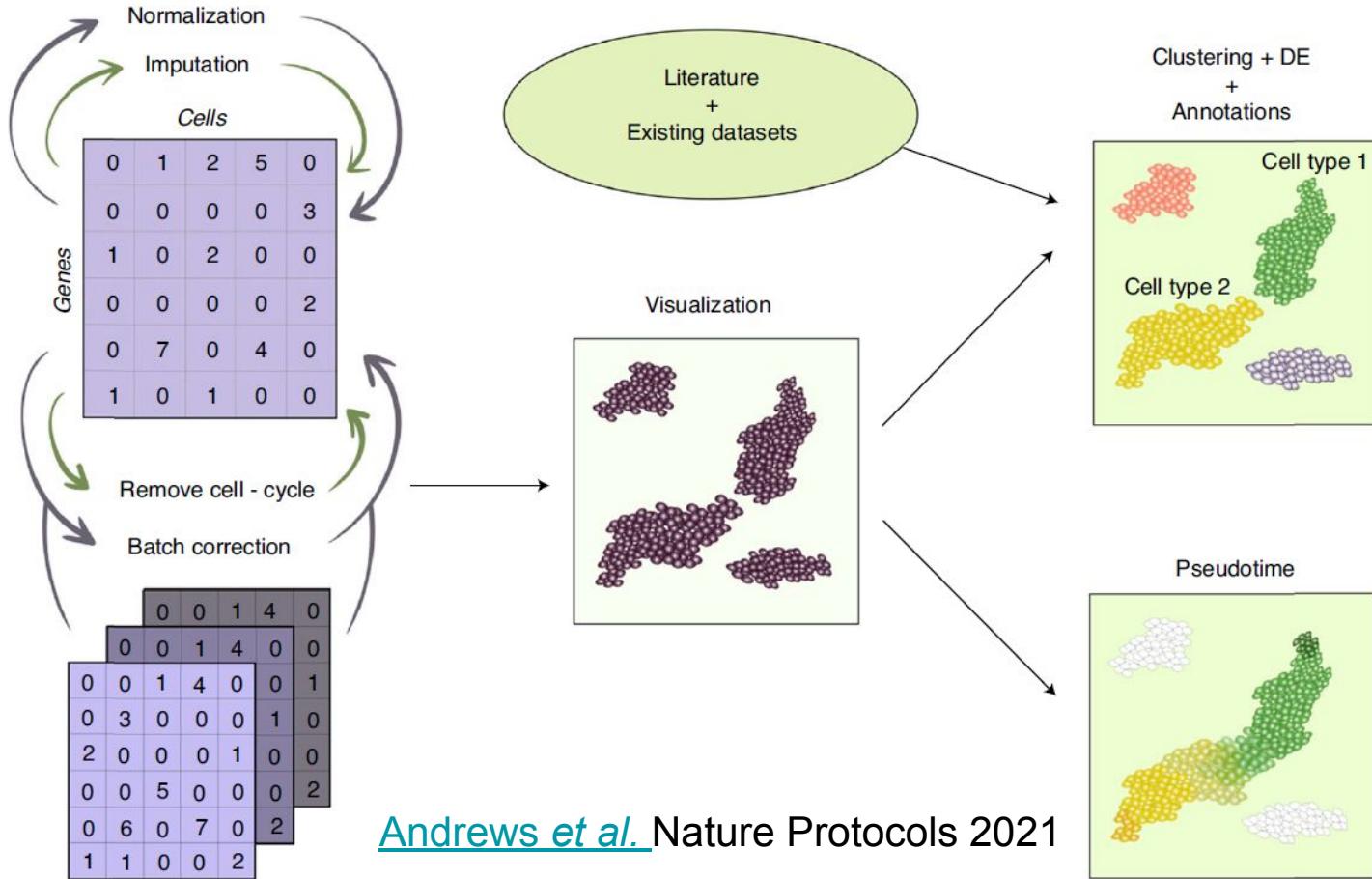
Marker genes



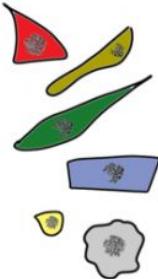
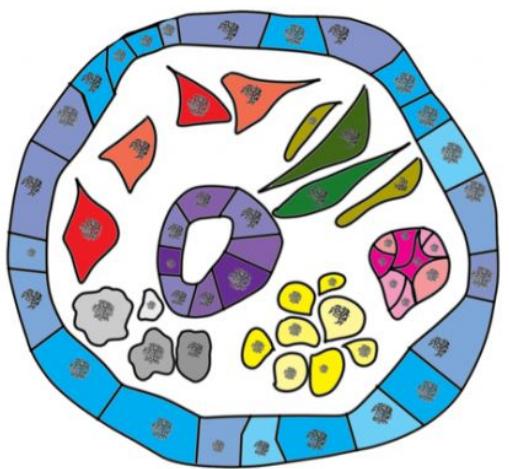
Expression patterns



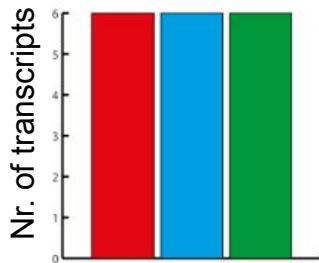
Overview of the computational workflow



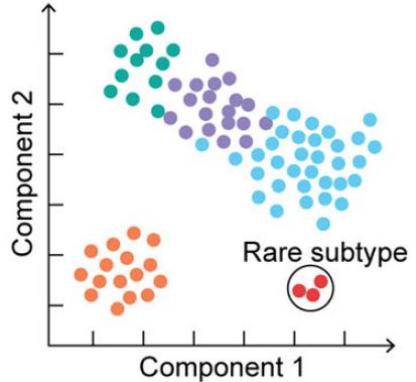
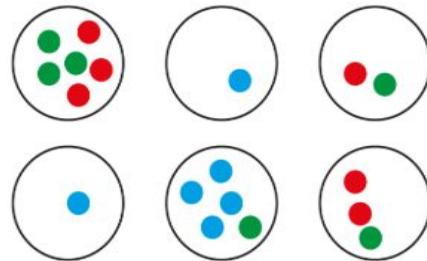
Single-cell biology benefits both disease understanding and drug discovery



Bulk analysis

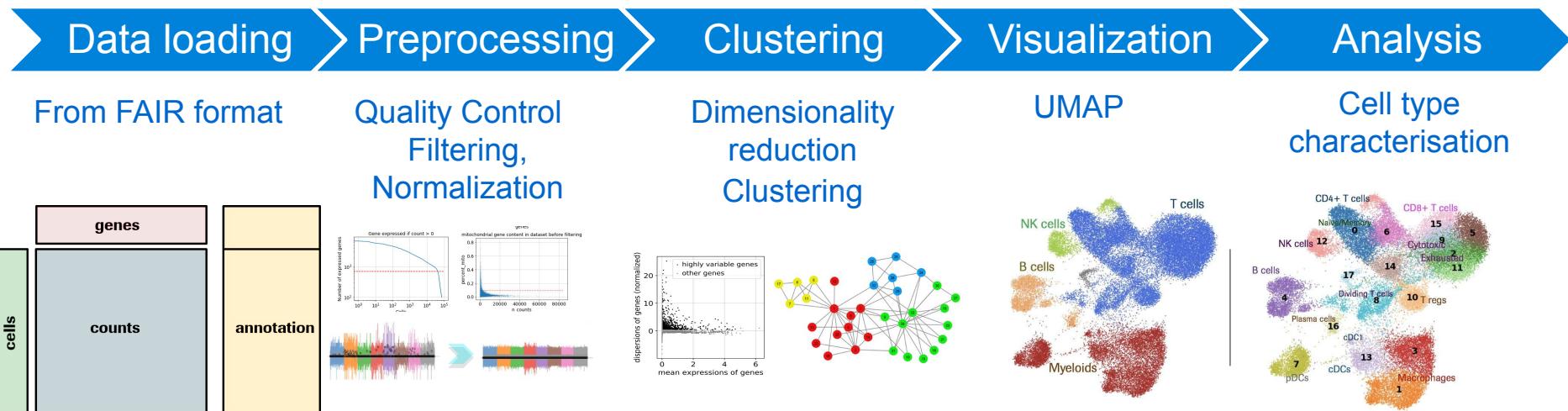


Single cell transcriptome analysis

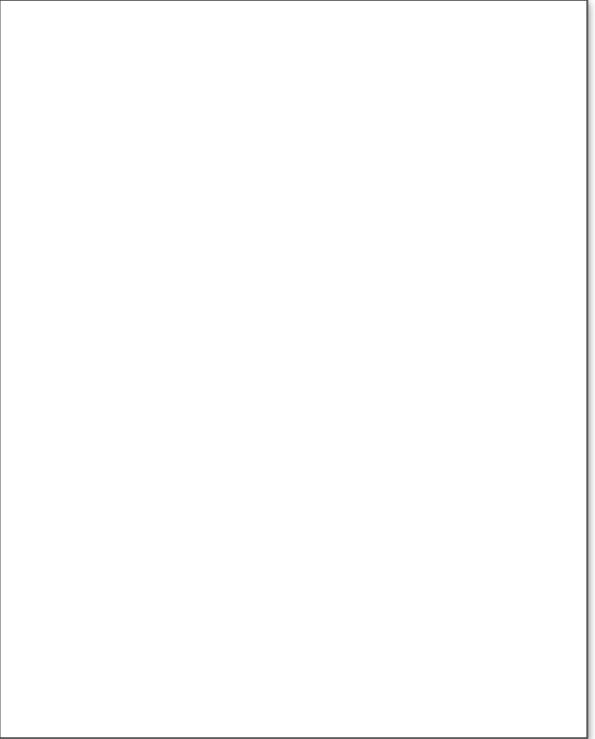


BESCA: An open-source Python package for single-cell gene expression analysis

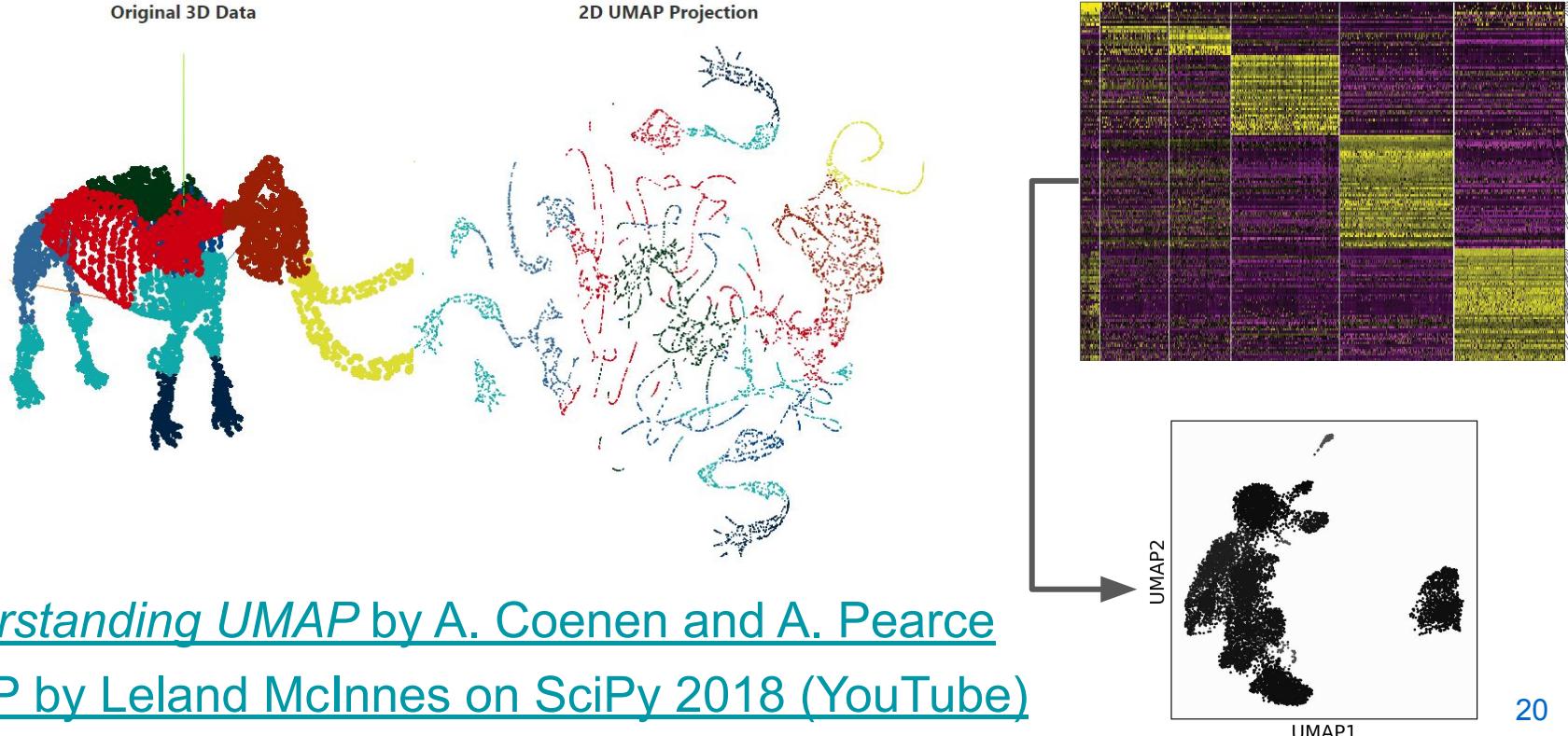
An automated standard workflow



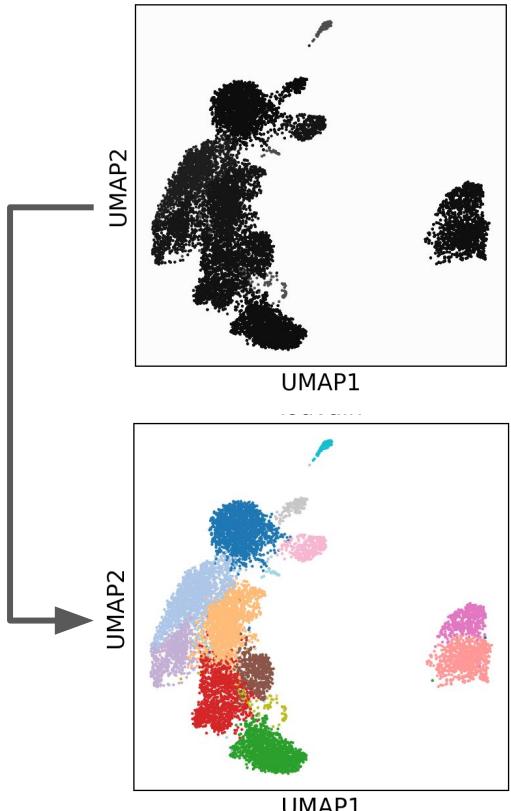
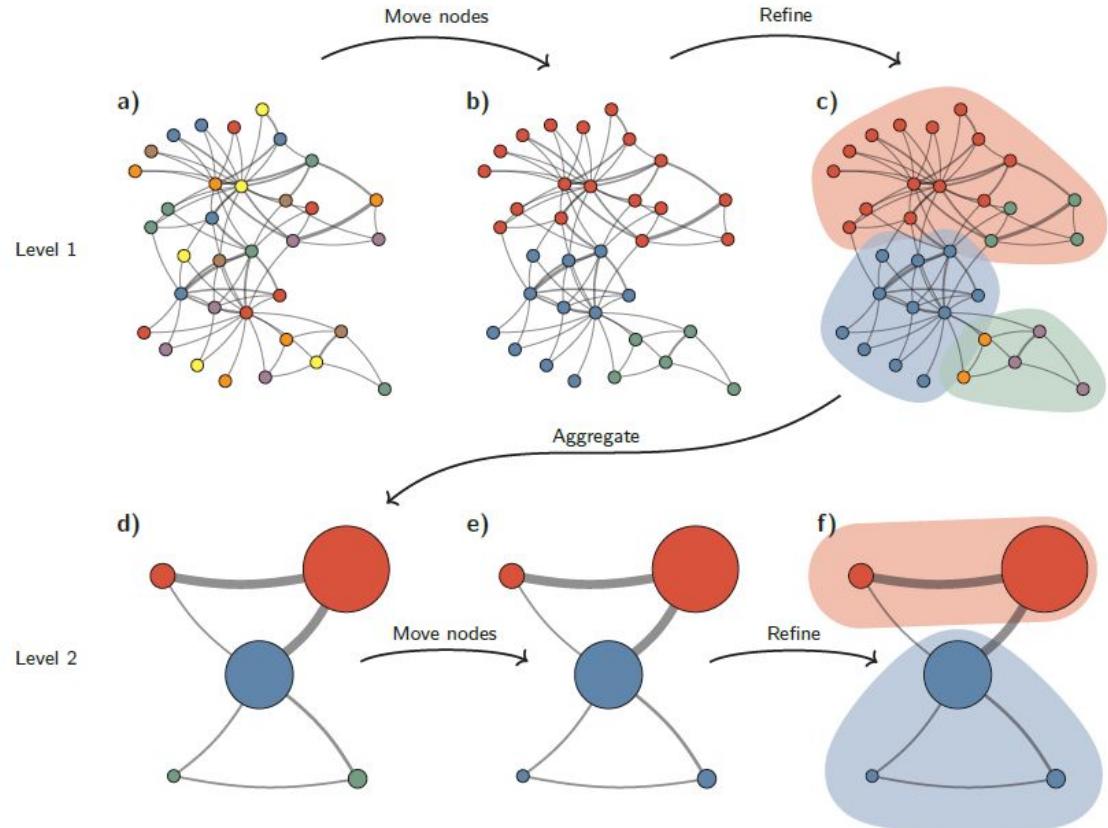
How to represent voxels with pixels?



Uniform Manifold Approximation and Projection (UMAP) for dimension reduction

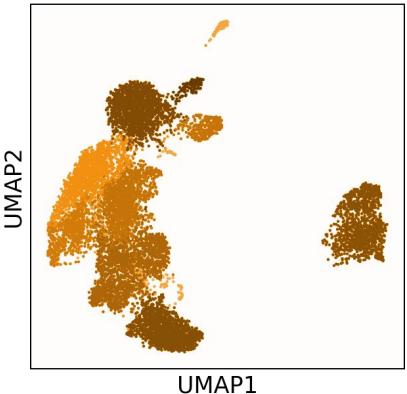


The Leiden Algorithm for Community Detection

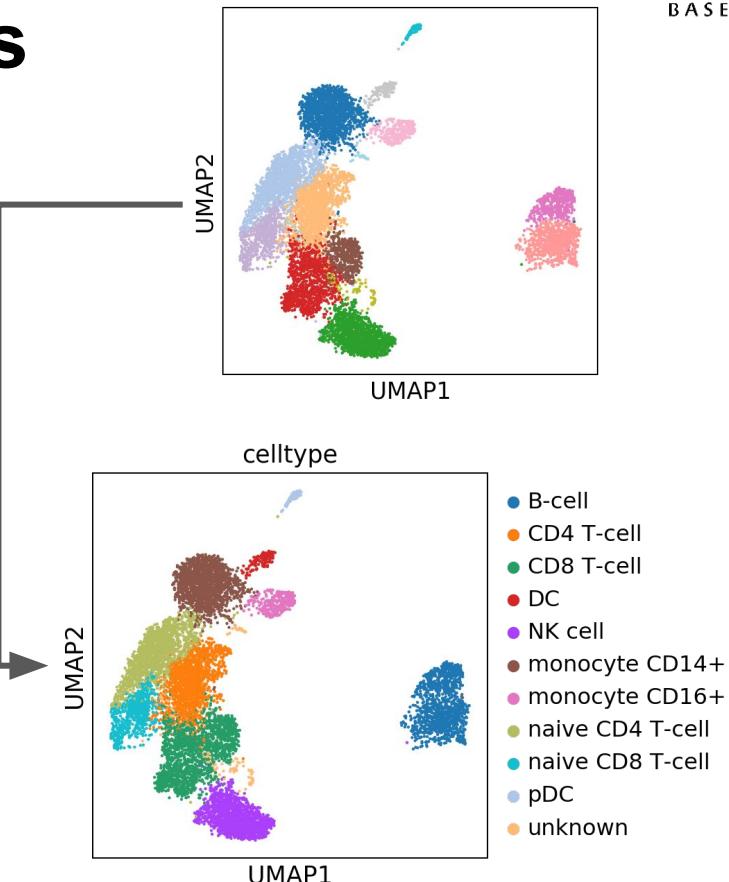


Biological knowledge and visual inspection is used to annotate cell types

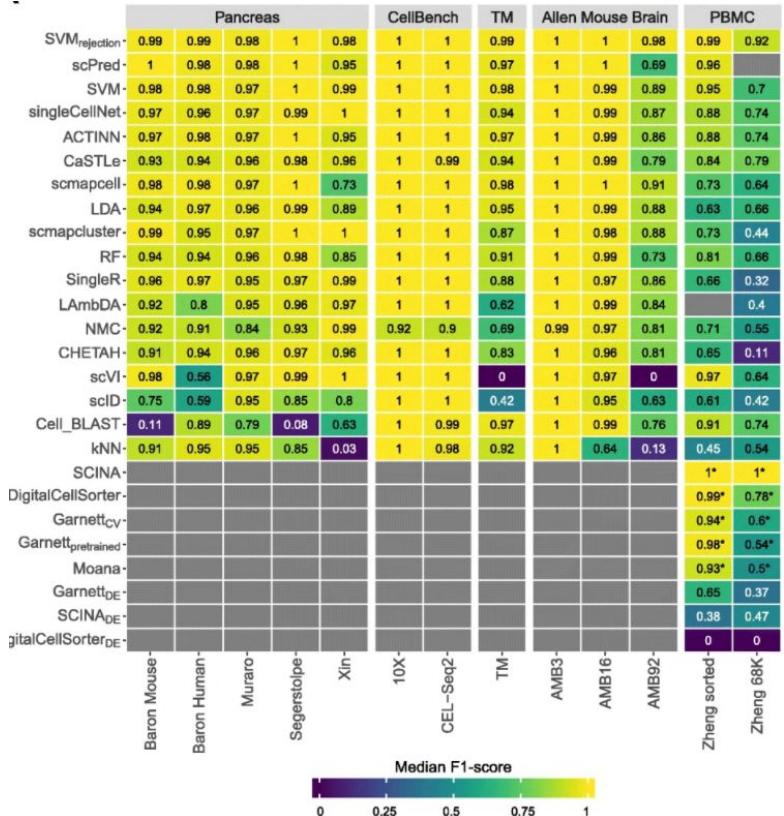
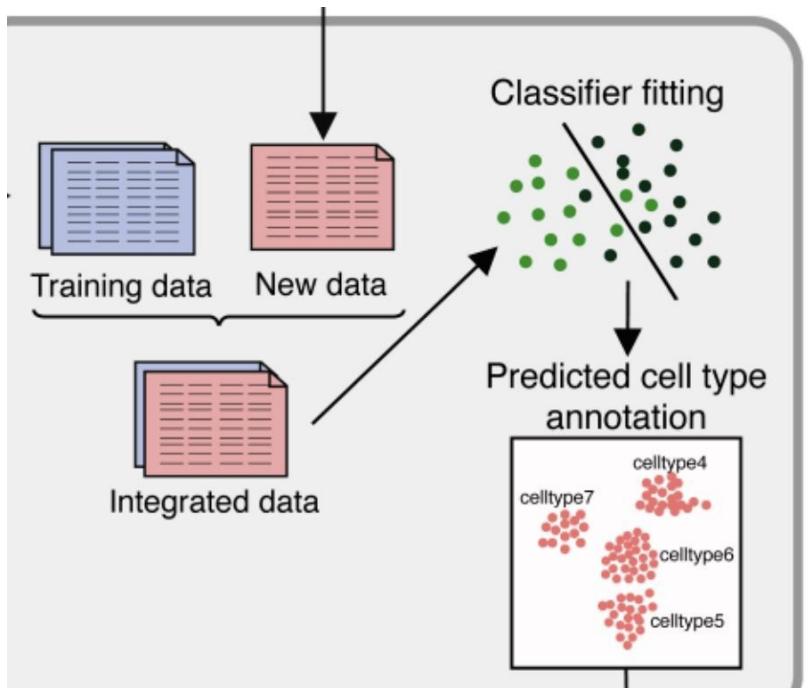
Heatmap
of gene X



lymphocyte	PTPRC							
myeloid	S100A8	S100A9	CST3					
Bcell	CD19	CD79A	MS4A1					
Tcells	CD3E	CD3G	CD3D					
CD4	CD4							
CD8	CD8A	CD8B						
NKcell	NKG7	GNLY	NCAM1					
monocyte	CST3	CSF1R	ITGAM	CD14	FCGR3A	FCGR3B		
macrophage	CD14	IL1B	LYZ	CD163	ITGAX	CD68	CSF1R	FCGR3A



Cell type annotation with machine learning

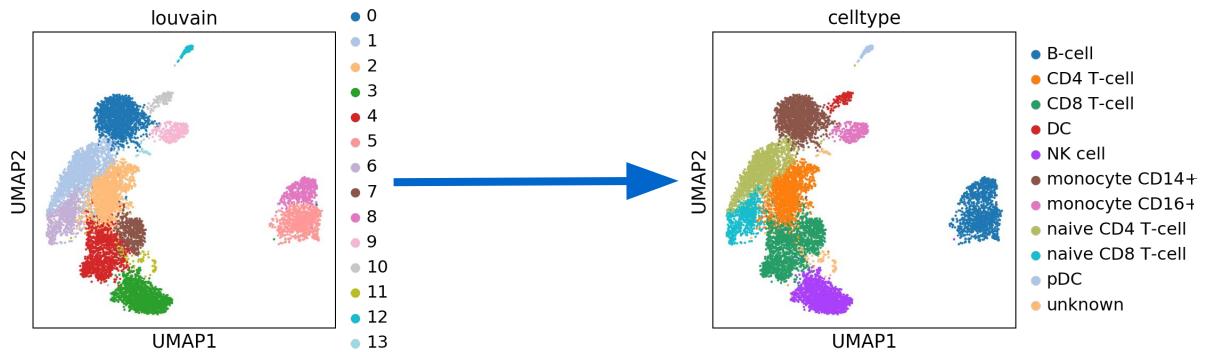


An intern project: Cell type annotation

From unsupervised clustering and cluster based annotation



Luis Wyss
RAAN intern 2019



	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5	Label
Training Cell 1	10	50	0	12	4	Celltype A
Training Cell 2	8	45	78	3	23	Celltype B
Training Cell 3	14	55	78	65	55	Celltype B
Training Cell 4	78	12	13	9	58	Celltype A
Training Cell 5	45	23	65	98	11	Celltype C

To supervised annotation at single-cell level:

	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5
Cell 1	45	45	8	56	3
Cell 2	65	120	78	45	12
Cell 3	79	12	34	65	88
Cell 4	7	59	32	47	62



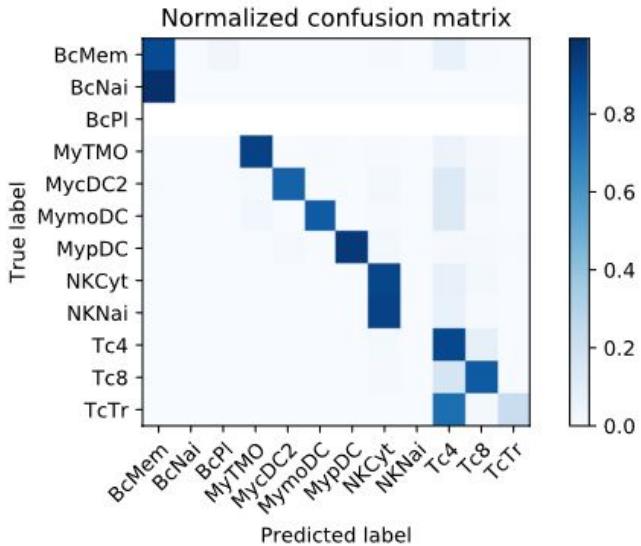
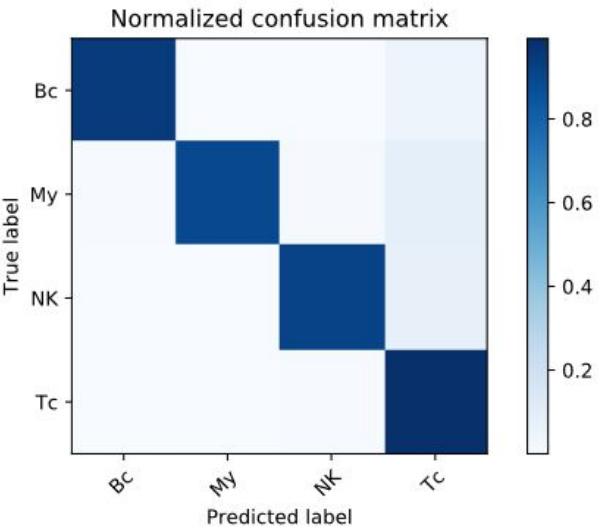
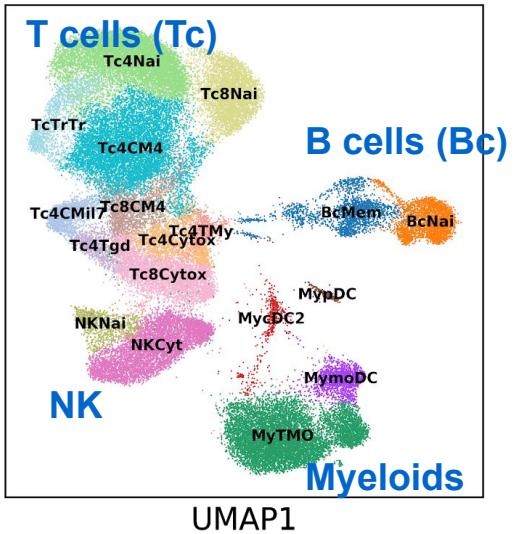
	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5	Prediction
Cell 1	45	45	8	56	3	Celltype A
Cell 2	65	120	78	45	12	Celltype B
Cell 3	79	12	34	65	88	Celltype C
Cell 4	7	59	32	47	62	Celltype B



Advantages: (1) automation, (2) annotation independent from clustering, and (3) we can estimate the confidence of prediction

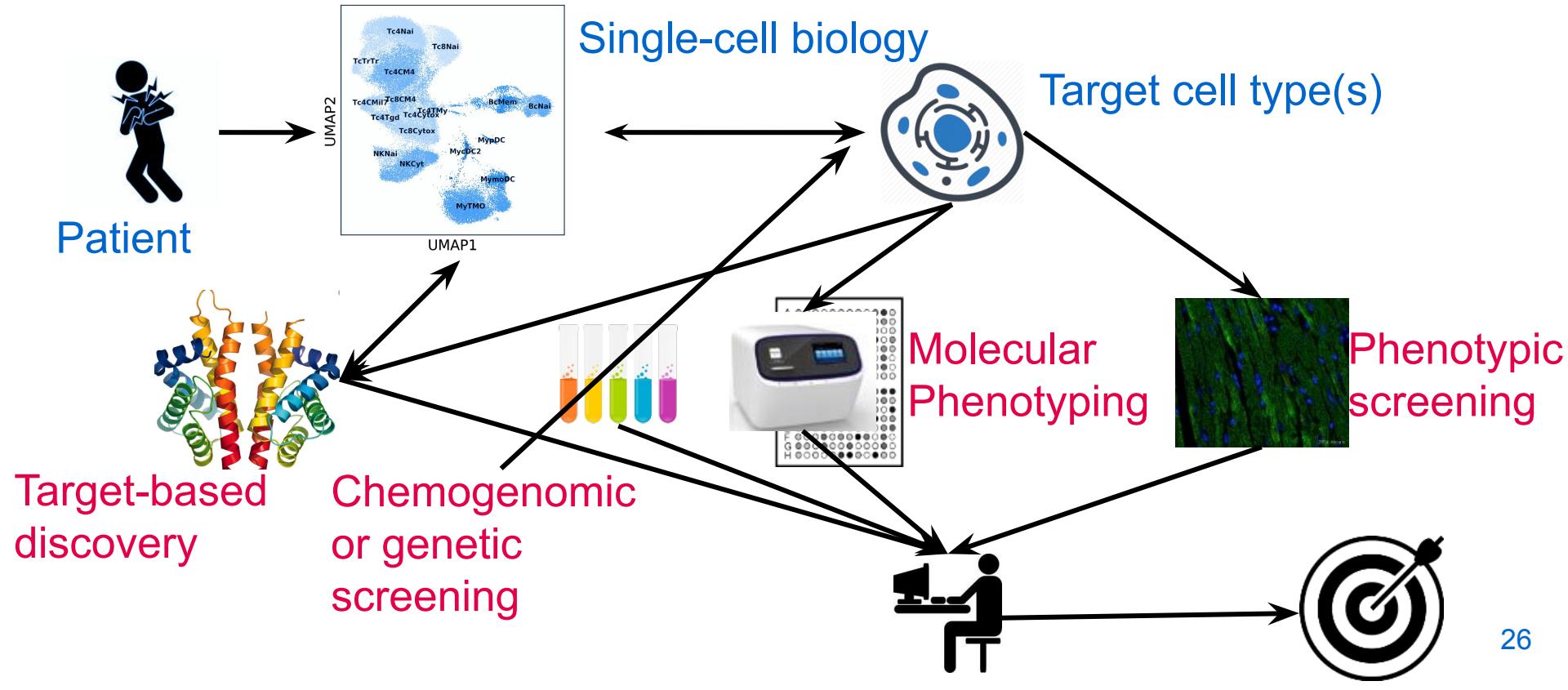
A PBMC example of cell type annotation

UMAP2



- Broad level cell types, including B cells (Bc), Myeloid (My), NK cells (NK) and T cells (Tc), are successfully predicted.
- Missing and highly similar cell types cause challenges with increased granularity. Essential: reference data quality and knowledge of cell types. 25

Computational biologists work with experimentalists to empower drug discovery



We are living ecosystems

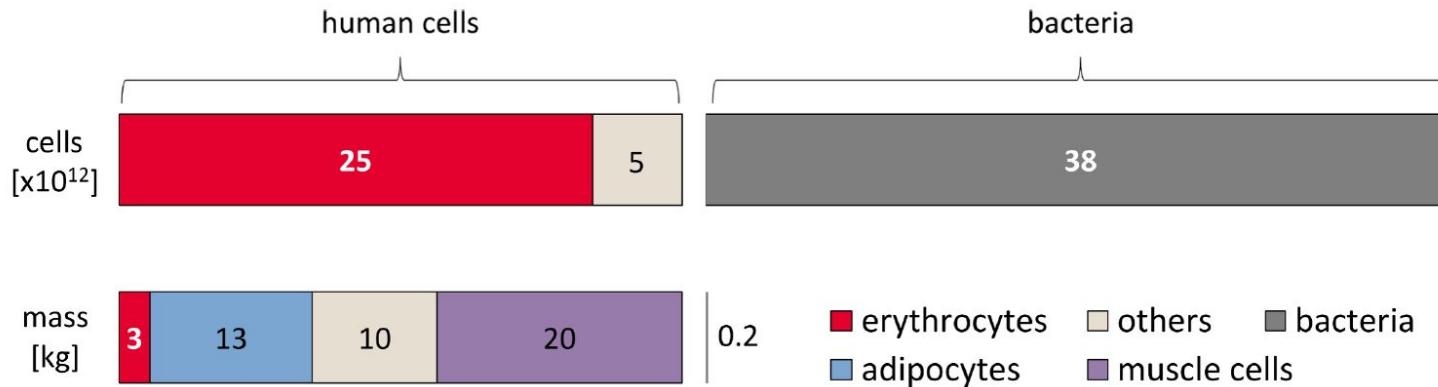
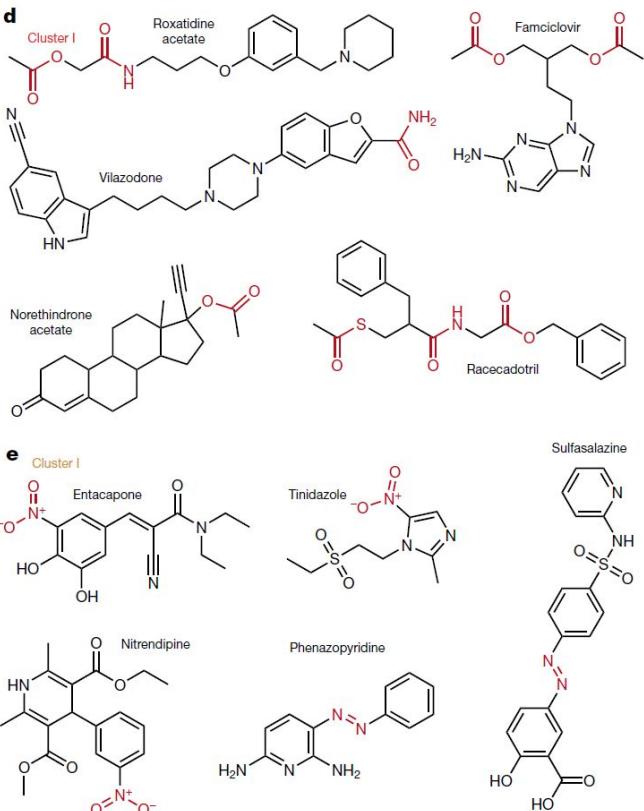
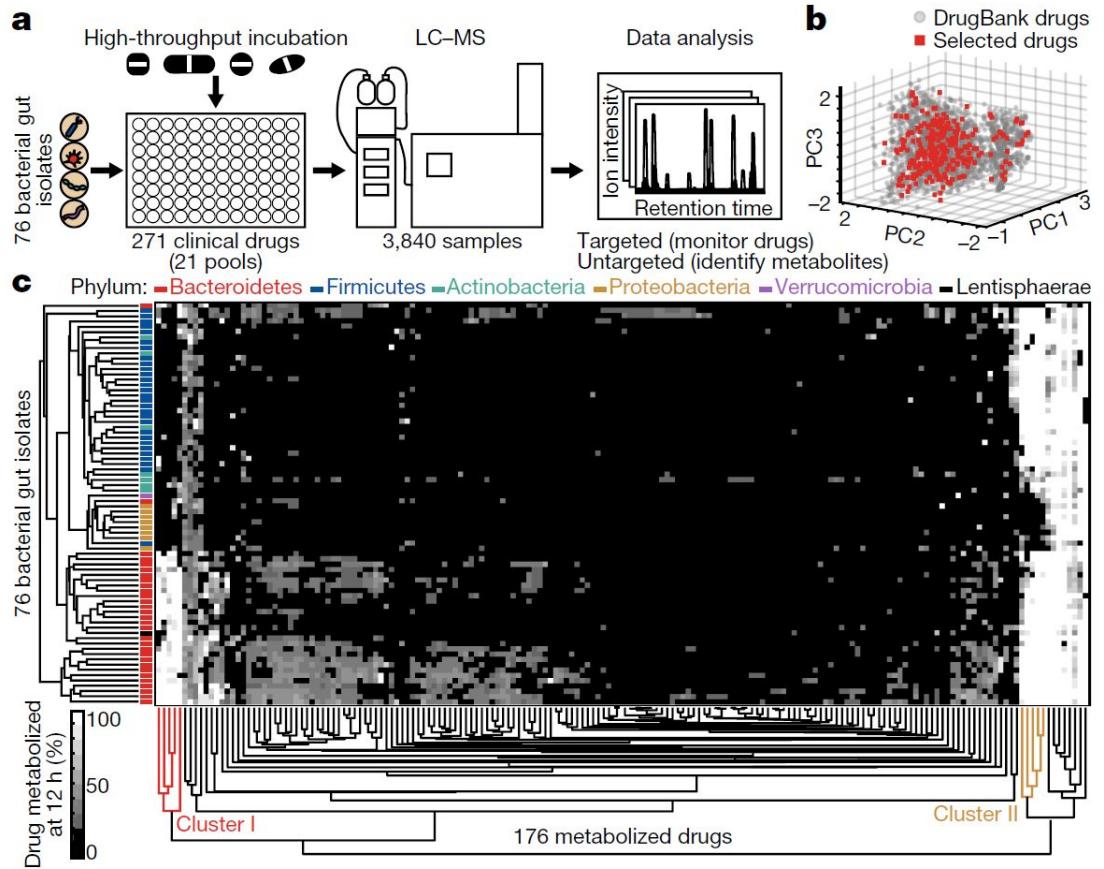


Table 3. B/H ratio for different population. See Table B in [S1 Appendix](#) for full references.

population segment	body weight [kg]	age [y]	blood volume [L]	RBC count $[10^{12}/L]$	colon content [g]	bac. conc. $[10^{11}/g \text{ wet}]^{(1)}$	total human cells $[10^{12}]^{(2)}$	total bacteria $[10^{12}]$	B:H
ref. man	70	20–30	4.9	5.0	420	0.92	30	38	1.3
ref. woman	63		3.9	4.5	480	0.92	21	44	2.2
young infant	4.4	4 weeks	0.4	3.8	48	0.92	1.9	4.4	2.3
infant	9.6	1	0.8	4.5	80	0.92	4	7	1.7
elder	70	66	3.8 ⁽³⁾	4.8	420	0.92	22	38	1.8
obese	140		6.7	5.0 ⁽⁴⁾	610 ⁽⁵⁾	0.92	40	56	1.4

Gut microbiome can metabolize drugs differently



Conclusions

- Single-cell biology can identify rare cell populations associated with diseases, and investigate cell-type-specific perturbations caused by drug candidates.
- Algorithms for dimensionality reduction, clustering, and semi-automated cell type annotation allow us interpret and integrate single-cell datasets.

Offline activities of Module IV (optional)

Perform your own single-cell data analysis to get first-hand experience working with high-dimensional biological data.

- If you are new to the topic, please use [the PBMC tutorial of Scanpy \(python\)](#) or [the PBMC tutorial of Seurat \(R\)](#).
- If you have experience with such data already, checkout [the NBIS workshop on single-cell sequencing data analysis](#) to cover advanced topics such as spatial transcriptomics and trajectory inference.

Single-cell biology is important in drug discovery

Disease understanding:

disease-specific cell types
and states



Target identification:

expression pattern in
health and disease across
cell types



Biomarker and patient stratification:

which genes should we measure
in which cell type(s)?



MoA and safety

modelling: perturbation effect at single-cell level



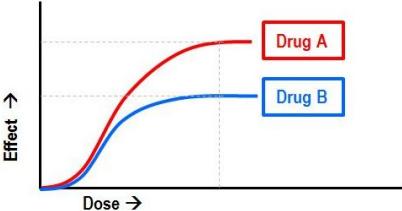
To understand protein function is critical for drug discovery



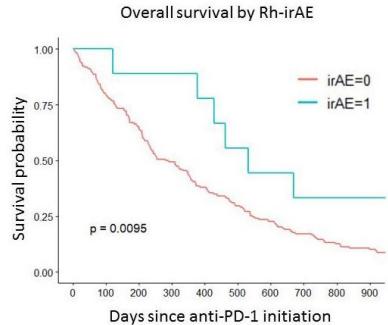
Protein function



Cellular phenotype



Drug efficacy and safety



Clinical outcome

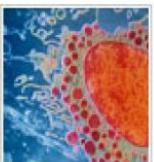
ONCOLOGY



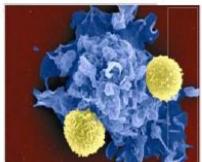
INFECTIOUS DISEASE



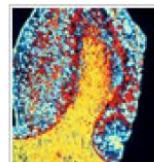
IMMUNOLOGY



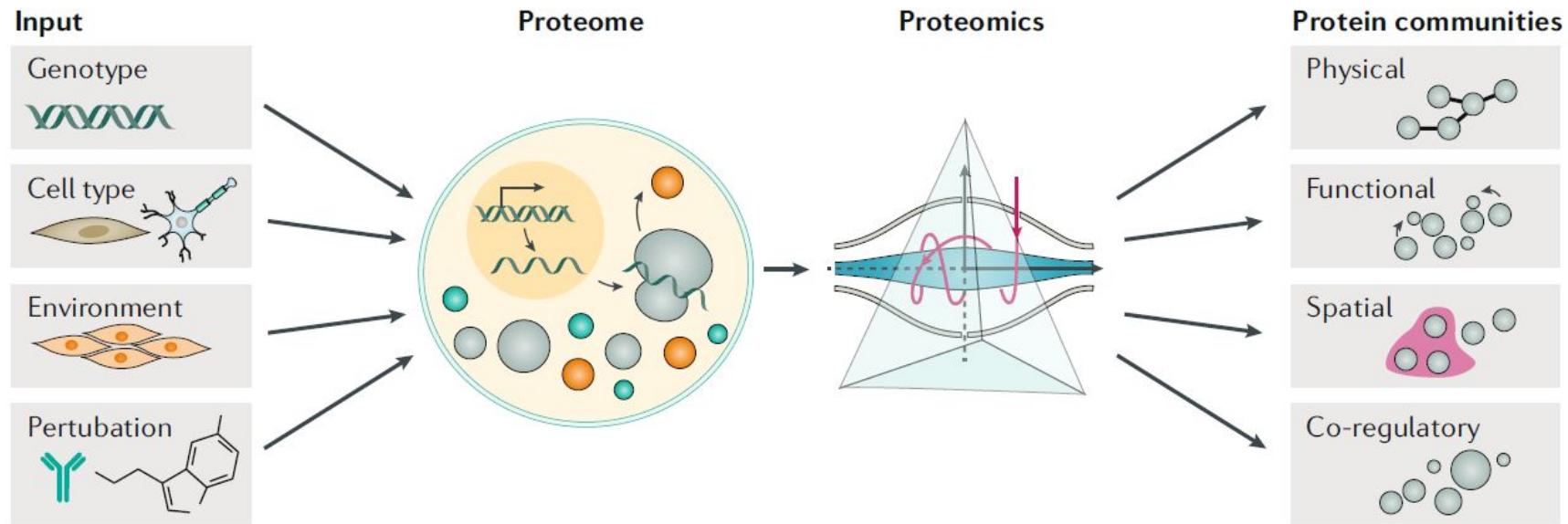
TUMOUR IMMUNOLOGY



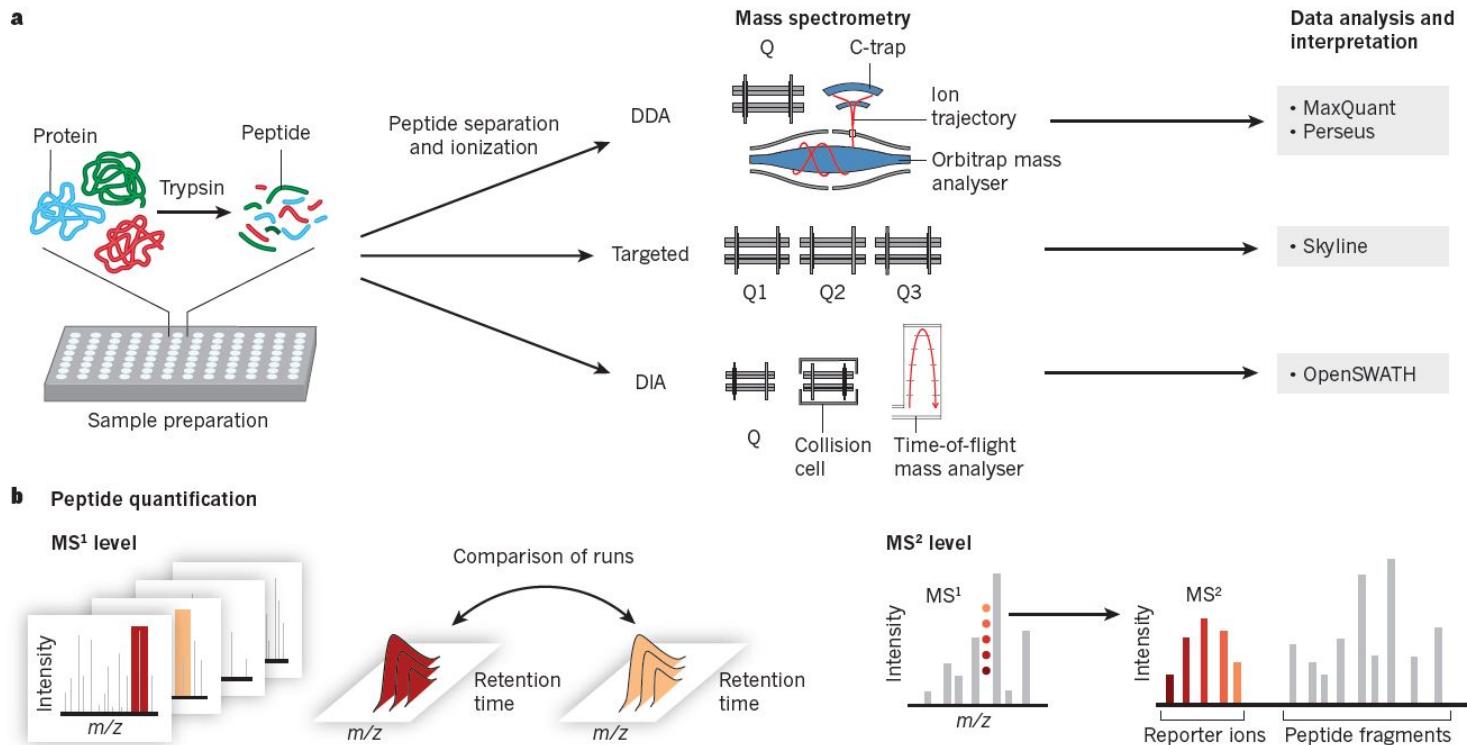
NEUROSCIENCES



Proteomics enables the elucidation of protein relations in the protein communities

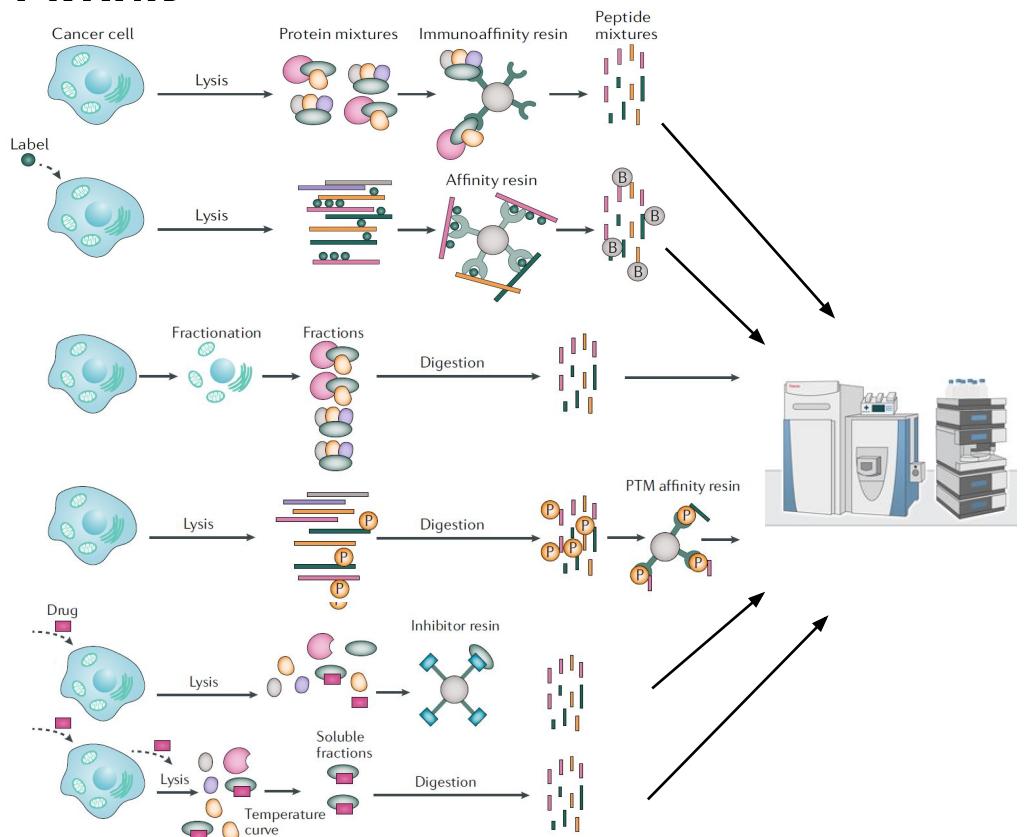


Advancement of proteomic technologies to characterize proteomes



Proteomics approaches for drug discovery

Applied for the characterization of protein communities and drug targets



Experimental approach

Affinity purification

Applications in drug discovery

Identification of drug-dependent **PPIs** and **protein-drug** interactions; characterization of **protein complex stoichiometry**

Proximity labelling

Characterization of **unbounded cellular compartments**, and **transient PPIs**

Organelle proteome profiling

Characterization of protein communities of all major **organelles** in one dataset, drug-induced **translocation**, and **secreted signaling factors**

PTM profiling

Characterization of **time- and dose-dependent signaling events** mediated by PTM

Chemoaffinity enrichment

Potential **drug targets** and **off-targets**

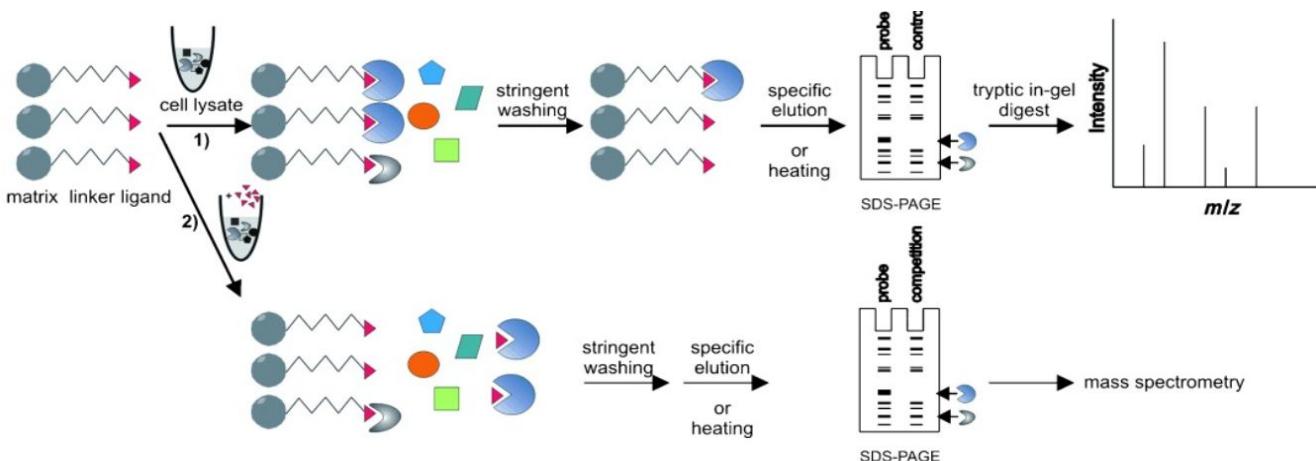
Thermal proteome profiling

Potential **drug targets** and **off-targets**

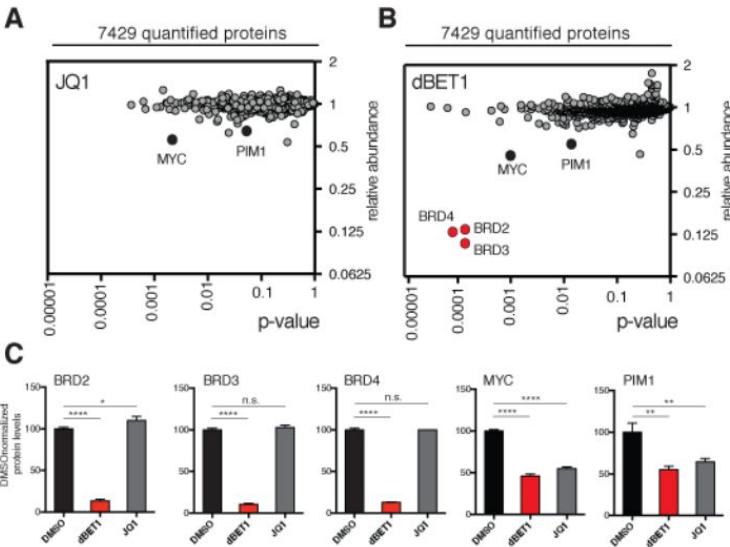
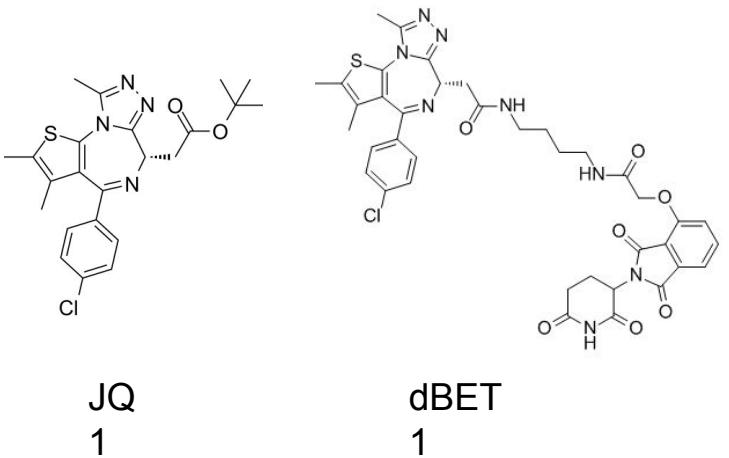
Example 1: Chemoproteomics methods for target ID



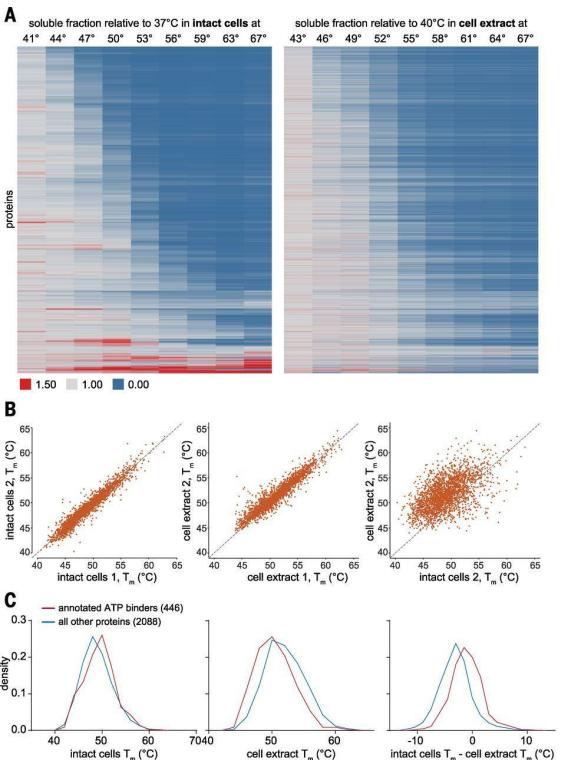
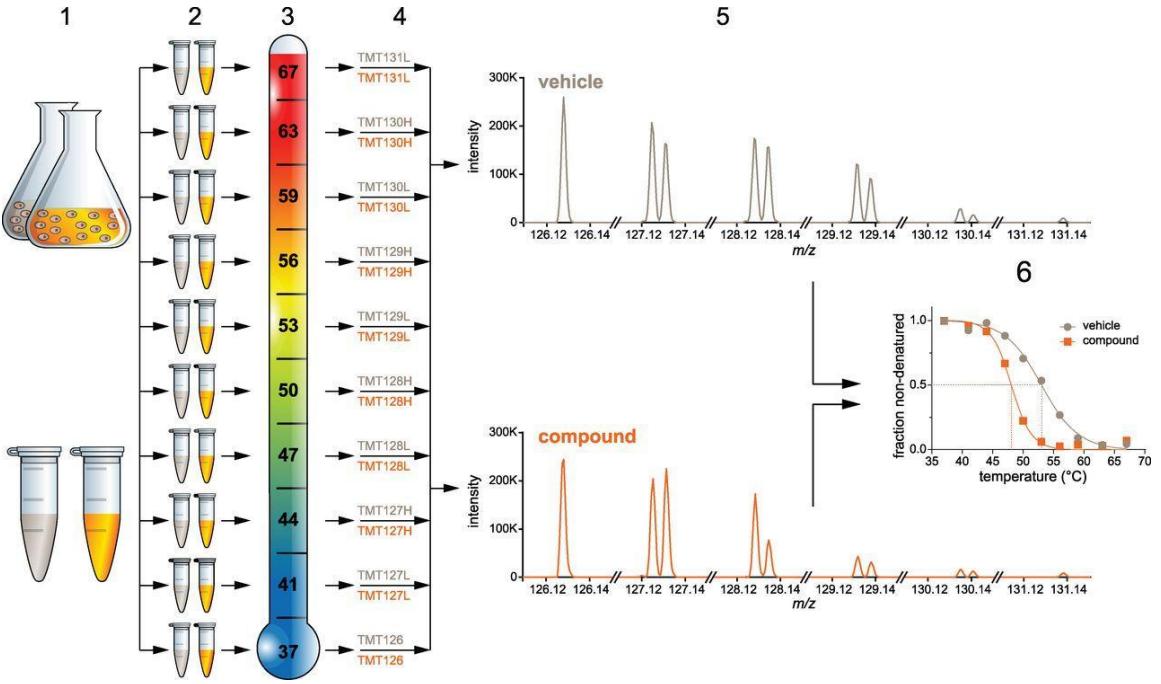
- Chemoproteomics methods are based on two principles:
 (1) **bait/prey** and (2) **competition**.
- Commonly used methods to identify binding partners of small molecules include affinity-based profiling (shown below), activity-based profiling, SILAC, etc.



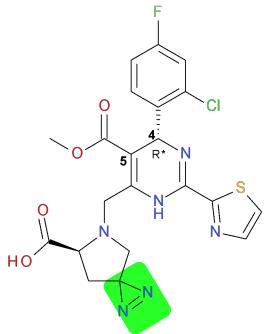
Example 2: Quantitative proteomics demonstrated selective degradation of proteins in a large set of protein mixtures



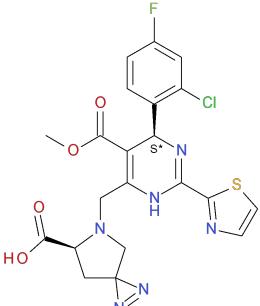
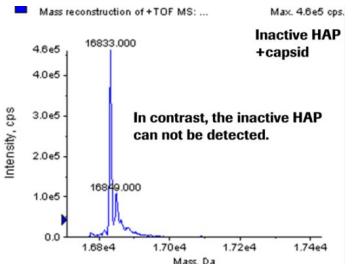
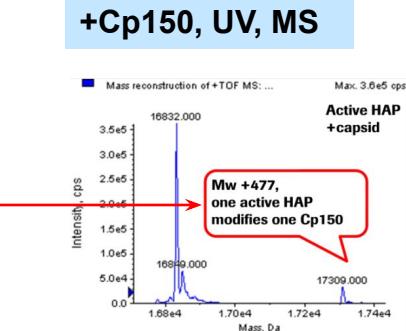
Example 3: thermal proteome profiling to track cancer drug in living cells (CETSA)



Example 4: photoaffinity labelling confirmed HBV capsid binding and mapped the small molecule binding pocket

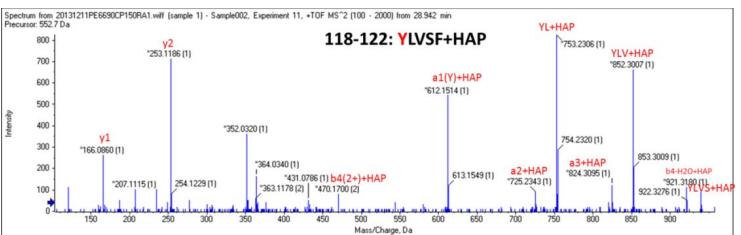


RO-A
 EC_{50} : **0.040 μM**
 IC_{50} : **0.47 μM**



RO-B
 EC_{50} : **>1 μM**
 IC_{50} : **>100 μM**

Proteolytic digestion/LC-MS/MS identified labelling site Y118



More photoaffinity labelling probes identified labelling sites at R127 and Y38.

Conclusion

- Mass spectrometry-based proteomics allows for target and off-target identification, the understanding of PPI, and biological pathways, time-, space- and dose-dependent signaling events etc
- Future directions could be to evaluate protein communities *in vivo*, to measure stoichiometry of proteins, and to provide methods for clinical samples.

References

1. Figures: [Lumen Learning](#), [Exploring Nature](#), [National Geographics](#), [Platelet cells](#) (Graham Beards, CC-BY-SA 4.0), [Lymphocytes](#) (Nicolas Grandjean, CC-BY-SA 3.0), [Adipocytes](#) (Public Domain), [Hepatocytes](#) (CC-BY-NC 2.0), [Neurons and Glia](#) (Public Domain), [Blood](#) (CC 3.0), [Blood Cells](#) (By A. Rad and M. Häggström. CC-BY-SA 3.0 license), [A selective JAK3 inhibitor](#) (London Lab/Weizmann institute)
2. Sender, Ron, Shai Fuchs, and Ron Milo. 2016. "Revised Estimates for the Number of Human and Bacteria Cells in the Body." PLoS Biology 14 (8). <https://doi.org/10.1371/journal.pbio.1002533>.
3. www.evocell-itn.eu;
4. Macaulay, Iain C., and Thierry Voet. 2014. "Single Cell Genomics: Advances and Future Perspectives." PLOS Genetics 10 (1): e1004126. <https://doi.org/10.1371/journal.pgen.1004126>.
5. Pryor, Rosina, Povilas Norvaisas, Georgios Marinos, Lena Best, Louise B. Thingholm, Leonor M. Quintaneiro, Wouter De Haes, et al. 2019. "Host-Microbe-Drug-Nutrient Screen Identifies Bacterial Effectors of Metformin Therapy." Cell 178 (6): 1299-1312.e29. <https://doi.org/10.1016/j.cell.2019.08.003>.
6. Cully, Megan. 2019. "Microbiome Therapeutics Go Small Molecule." *Nature Reviews Drug Discovery* 18 (July): 569. <https://doi.org/10.1038/d41573-019-00122-8>.
7. Duscha, Alexander, Barbara Gisevius, Sarah Hirschberg, Nissan Yissachar, Gabriele I. Stangl, Eva Eilers, Verian Bader, et al. 2020. "Propionic Acid Shapes the Multiple Sclerosis Disease Course by an Immunomodulatory Mechanism." Cell 180 (6): 1067-1080.e16. <https://doi.org/10.1016/j.cell.2020.02.035>.
8. Pryor, Rosina, Povilas Norvaisas, Georgios Marinos, Lena Best, Louise B. Thingholm, Leonor M. Quintaneiro, Wouter De Haes, et al. 2019. "Host-Microbe-Drug-Nutrient Screen Identifies Bacterial Effectors of Metformin Therapy." Cell 178 (6): 1299-1312.e29. <https://doi.org/10.1016/j.cell.2019.08.003>.
9. Zimmermann, Michael, Maria Zimmermann-Kogadeeva, Rebekka Wegmann, and Andrew L. Goodman. 2019. "Mapping Human Microbiome Drug Metabolism by Gut Bacteria and Their Genes." *Nature* 570 (7762): 462. <https://doi.org/10.1038/s41586-019-1291-3>.
10. Shin, Hyun Kil, Young-Mook Kang, and Kyoung Tai No. 2016. "Predicting ADME Properties of Chemicals." In *Handbook of Computational Chemistry*, edited by Jerzy Leszczynski, 1–37. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6169-8_59-1.

References (continued)

11. Mädler, Sophia Clara, Alice Julien-Laferriere, Luis Wyss, Miroslav Phan, Albert S. W. Kang, Eric Ulrich, Roland Schmucki, et al. 2020. "Besca, a Single-Cell Transcriptomics Analysis Toolkit to Accelerate Translational Research." *BioRxiv*, September, 2020.08.11.245795. <https://doi.org/10.1101/2020.08.11.245795>.
12. Andrews, Tallulah S., Vladimir Yu Kiselev, Davis McCarthy, and Martin Hemberg. 2021. "Tutorial: Guidelines for the Computational Analysis of Single-Cell RNA Sequencing Data." *Nature Protocols* 16 (1): 1–9. <https://doi.org/10.1038/s41596-020-00409-w>.
13. Sturm, Gregor, Francesca Finotello, Florent Petitprez, Jitao David Zhang, Jan Baumbach, Wolf H. Fridman, Markus List, and Tatsiana Aneichyk. 2019. "Comprehensive Evaluation of Transcriptome-Based Cell-Type Quantification Methods for Immuno-Oncology." *Bioinformatics* 35 (14): i436–45. <https://doi.org/10.1093/bioinformatics/btz363>.
14. Villani, Alexandra-Chloé, Rahul Satija, Gary Reynolds, Siranush Sarkizova, Karthik Shekhar, James Fletcher, Morgane Griesbeck, et al. 2017. "Single-Cell RNA-Seq Reveals New Types of Human Blood Dendritic Cells, Monocytes, and Progenitors." *Science* 356 (6335): eaah4573. <https://doi.org/10.1126/science.aah4573>.
15. Finotello, Francesca, Clemens Mayer, Christina Plattner, Gerhard Laschober, Dietmar Rieder, Hubert Hackl, Anne Krogsdam, et al. 2019. "Molecular and Pharmacological Modulators of the Tumor Immune Contexture Revealed by Deconvolution of RNA-Seq Data." *Genome Medicine* 11 (1): 34. <https://doi.org/10.1186/s13073-019-0638-6>.
16. Fridman, Wolf H., Laurence Zitvogel, Catherine Sautès-Fridman, and Guido Kroemer. 2017. "The Immune Contexture in Cancer Prognosis and Treatment." *Nature Reviews Clinical Oncology* 14 (12): 717–34. <https://doi.org/10.1038/nrclinonc.2017.101>.
17. Moisan, Annie, Marcel Gubler, Jitao David Zhang, Yann Tessier, Kamille Dumong Erichsen, Sabine Sewing, Régine Gérard, et al. 2017. "Inhibition of EGF Uptake by Nephrotoxic Antisense Drugs In Vitro and Implications for Preclinical Safety Profiling." *Molecular Therapy - Nucleic Acids* 6 (March): 89–105. <https://doi.org/10.1016/j.omtn.2016.11.006>.
18. Chang, Chia-Yu, Hsiao-Chien Ting, Ching-Ann Liu, Hong-Lin Su, Tzyy-Wen Chiou, Horng-Jyh Harn, and Shinn-Zong Lin. 2018. "Induced Pluripotent Stem Cells: A Powerful Neurodegenerative Disease Modeling Tool for Mechanism Study and Drug Discovery." *Cell Transplantation* 27 (June): 096368971877540. <https://doi.org/10.1177/0963689718775406>.

References (continued)

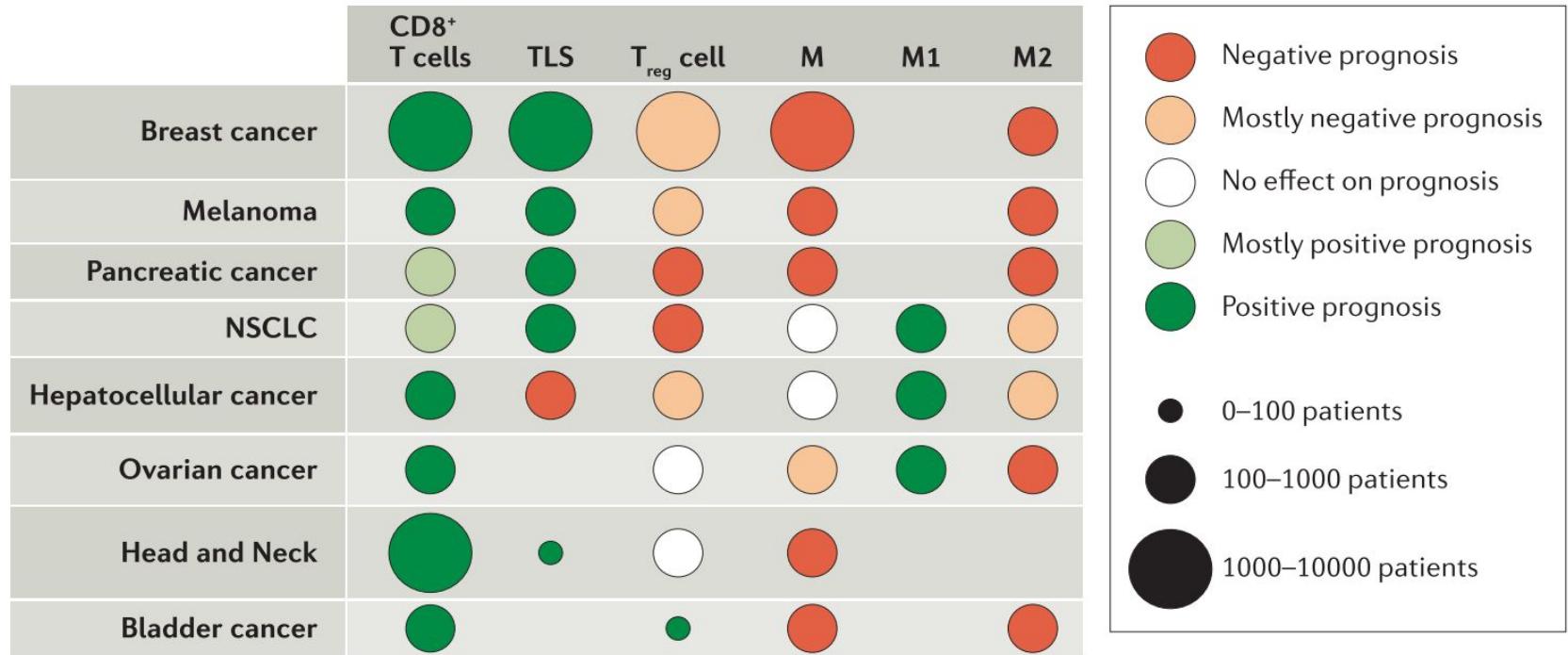
19. Takahashi, Toshio. 2019. "Organoids for Drug Discovery and Personalized Medicine." *Annual Review of Pharmacology and Toxicology* 59 (1): 447–62. <https://doi.org/10.1146/annurev-pharmtox-010818-021108>.
20. Budayeva, Hanna G., and Donald S. Kirkpatrick. 2020. "Monitoring Protein Communities and Their Responses to Therapeutics." *Nature Reviews Drug Discovery* 19 (6): 414–26. <https://doi.org/10.1038/s41573-020-0063-y>.
21. Lukonin, Ilya, Denise Serra, Ludivine Challet Meylan, Katrin Volkmann, Janine Baaten, Rui Zhao, Shelly Meeusen, et al. 2020. "Phenotypic Landscape of Intestinal Organoid Regeneration." *Nature* 586 (7828): 275–80. <https://doi.org/10.1038/s41586-020-2776-9>.
22. Drawnel, Faye M., Stefano Boccardo, Michael Prummer, Frédéric Delobel, Alexandra Graff, Michael Weber, Régine Gérard, et al. 2014. "Disease Modeling and Phenotypic Drug Screening for Diabetic Cardiomyopathy Using Human Induced Pluripotent Stem Cells." *Cell Reports* 9 (3): 810–20. <https://doi.org/10.1016/j.celrep.2014.09.055>.
23. Traag, Vincent, Ludo Waltman, and Nees Jan van Eck. 2019. "From Louvain to Leiden: Guaranteeing Well-Connected Communities." *Scientific Reports* 9 (1): 5233. <https://doi.org/10.1038/s41598-019-41695-z>.
24. *Understanding UMAP*, Andy Coenen and Adam Pearce, <https://pair-code.github.io/understanding-umap/>
25. How exactly UMAP works, Nikolay Oskolkov, <https://towardsdatascience.com/how-exactly-umap-works-13e3040e1668>
26. McInnes, Leland, and John Healy. 2018. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." ArXiv:1802.03426 [Cs, Stat], February. <http://arxiv.org/abs/1802.03426>.
27. Zappia, Luke, Belinda Phipson, and Alicia Oshlack. 2018. "Exploring the Single-Cell RNA-Seq Analysis Landscape with the ScRNA-Tools Database." *PLOS Computational Biology* 14 (6): e1006245. <https://doi.org/10.1371/journal.pcbi.1006245>.
28. Abdelaal, Tamim, Lieke Michielsen, Davy Cats, Dylan Hoogduin, Hailiang Mei, Marcel J. T. Reinders, and Ahmed Mahfouz. 2019. "A Comparison of Automatic Cell Identification Methods for Single-Cell RNA Sequencing Data." *Genome Biology* 20 (1): 194. <https://doi.org/10.1186/s13059-019-1795-z>.
- 29.
- 30.

Supplementary Information

Embryonic origins of tissues

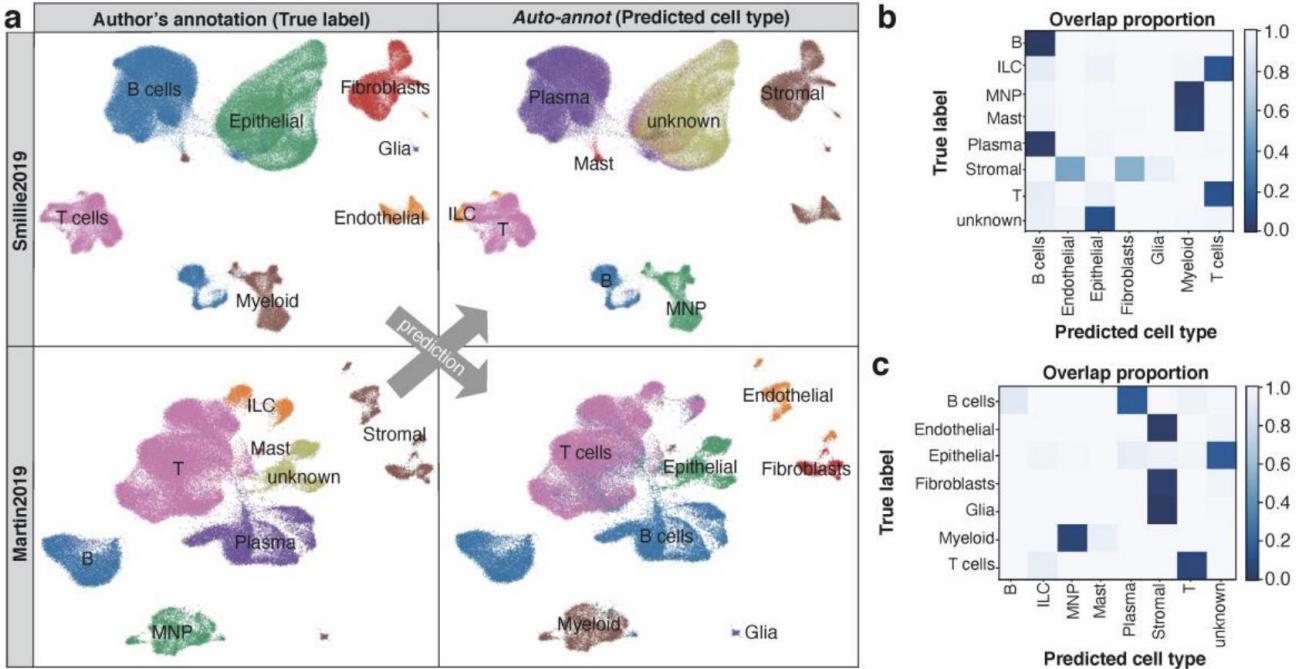
Germ Layer	Gives rise to:		
Ectoderm	Epidermis, glands on skin, some cranial bones, pituitary and adrenal medulla, the nervous system, the mouth between cheek and gums, the anus		
Mesoderm	Connective tissues proper, bone, cartilage, blood, endothelium of blood vessels, muscle, synovial membranes, serous membranes lining body cavities, kidneys, lining of gonads		
Endoderm	Lining of airways and digestive system except the mouth and distal part of digestive system (rectum and anal canal); glands (digestive glands, endocrine glands, adrenal cortex)		

Abundance of immune cells in tumor microenvironments affect outcome



TLS: tertiary lymphoid structures; T_{reg}: regulatory T cells; M: macrophages; M1/M2: subtypes of macrophages

An example of Inflammatory Bowel Disease (IBD)



We observed Inconsistent cell type nomenclature across studies.
Machine learning allows us compare and integrate multiple studies.